

Db	181	GGCCACCAAGATAAAGGACTTGCACCGAGCGCCAGGCCACTTCTTCCGCGAGAACCTGAGCC	240
Qy	241	TTCCCCCAAGGGCAAGGCCCGCGAGTTCCCCAGCGACAGAACCCGCGCAACAGCCCCACC	300
Db	241	TTCCCCCAAGGGCAAGGCCCGCGAGTTCCCCAGCGACAGAACCCGCGCAACAGCCCCACC	300
Qy	301	AGCCCGCAGCTGCAAGTGTGCGCGCGCAACCCCGCAGCGAGGCCGCGCGCTGAGGCCCAAG	360
Db	301	AGCCCGCAGCTGCAAGTGTGCGCGCGCAACCCCGCAGCGAGGCCGCGCGCTGAGGCCCAAG	360
Qy	361	GGCAACCTGAACTTCCCGCCAGATCAACCTGTGGCAGCGCCCCCTGTGAGCATCAAGGTG	420
Db	361	GGCAACCTGAACTTCCCGCCAGATCAACCTGTGGCAGCGCCCCCTGTGAGCATCAAGGTG	420
Qy	421	GGCGGCGCAGATCAAGAGGCGCCCTGCTGACAACCGCGCGCGACACACCTGTGAGAGAG	480
Db	421	GGCGGCGCAGATCAAGAGGCGCCCTGCTGACAACCGCGCGCGACACACCTGTGAGAGAG	480
Qy	481	ATGAGCTGTGCCCCCGGCAATGGAAGCCCAAGATATGCGCGGCATTCGCGGCTTCATCAAG	540
Db	481	ATGAGCTGTGCCCCCGGCAATGGAAGCCCAAGATATGCGCGGCATTCGCGGCTTCATCAAG	540
Qy	541	GTGGGCGCATGAGACGAGATCCGATGTGAGATCTGTGGCGCAAGAGGCCATTCGGCACCTG	600
Db	541	GTGGGCGCATGAGACGAGATCCGATGTGAGATCTGTGGCGCAAGAGGCCATTCGGCACCTG	600
Qy	601	CTGATCGGCCCCCAACCCCTGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGAGCTGC	660
Db	601	CTGATCGGCCCCCAACCCCTGTGAACATCATTCGCGCGCAACATGCTGACCCAGCTGAGCTGC	660
Qy	661	ACCCTGAACCTTCCCATCATGCCCCCATCGAGACCTGTGCGGTGAAGCTGAAGCCCGGCAATG	720
Db	661	ACCCTGAACCTTCCCATCATGCCCCCATCGAGACCTGTGCGGTGAAGCTGAAGCCCGGCAATG	720
Qy	721	GACGCGCCCCCAAGGTAAAGAGATGCGCCCTGAACCGAAGAAAGATTCAAAGGCGCTGACCGCC	780
Db	721	GACGCGCCCCCAAGGTAAAGAGATGCGCCCTGAACCGAAGAAAGATTCAAAGGCGCTGACCGCC	780
Qy	781	ATCTCGAGAGATATGAGAGAGGCGCAAGATCAACAAAGATCGGCCCCGAGAACCCCTTAC	840
Db	781	ATCTCGAGAGATATGAGAGAGGCGCAAGATCAACAAAGATCGGCCCCGAGAACCCCTTAC	840
Qy	841	AAACACCCCGCTGTTCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGTGTGAC	900
Db	841	AAACACCCCGCTGTTCGCCATCAAGAAAGAGACAGCAACCAAGTGGCGCAAGCTGTGTGAC	900
Qy	901	TTCCGCGAGCTGAACAAGGCGACCCAGAACTTGTGGAAGGTGACGTGTGGCATTCGCCAC	960
Db	901	TTCCGCGAGCTGAACAAGGCGACCCAGAACTTGTGGAAGGTGACGTGTGGCATTCGCCAC	960
Qy	961	CCCGCGGCGCTGAAGAAAGAGACGCTGACGCTGTGAGCGTGGGCGACGCGCTACTTC	1020
Db	961	CCCGCGGCGCTGAAGAAAGAGACGCTGACGCTGTGAGCGTGGGCGACGCGCTACTTC	1020
Qy	1021	AGCGTGCCTCTGAGCGAGGACTTTCGCGAAGTACACGCGCTTTCACATCCCAAGATCAAC	1080
Db	1021	AGCGTGCCTCTGAGCGAGGACTTTCGCGAAGTACACGCGCTTTCACATCCCAAGATCAAC	1080
Qy	1081	AAACGAGACCCCGGCGCATTCGCTACCAAGTTCMAAGTGTGCCCCAGGCTGTGAAGGGCAGC	1140
Db	1081	AAACGAGACCCCGGCGCATTCGCTACCAAGTTCMAAGTGTGCCCCAGGCTGTGAAGGGCAGC	1140
Qy	1141	CCCAAGATCTTCCAGAGCAGACGACCAAGATCTGTGAAGCCCTTCGCGCGCGCGCAACCCC	1200
Db	1141	CCCAAGATCTTCCAGAGCAGACGACCAAGATCTGTGAAGCCCTTCGCGCGCGCGCAACCCC	1200
Qy	1201	GAGATCGTATCTAACAGTACATGAGCGACCTGTATGTTGTTGGCAGCGACCTGTGAGATGTGGC	1260
Db	1201	GAGATCGTATCTAACAGTACATGAGCGACCTGTATGTTGTTGGCAGCGACCTGTGAGATGTGGC	1260
Qy	1261	CAGCACCGGCGCAAGATCGAGAGCTGTGCGCAGACACTGTGTGCTGTGGGCTTTCACCAAC	1320

Dd	1261	CAGCACC	CGCCAAAGATCGAGAGACCTGCGCAGCAGCACTGCTGCGCTGAGGCTTCACCAAC	1320
Qy	1321	CCCGCAAGAA	GACACAGAAAGAGCCCCCTCTCTGTGATGAGCTGCACCCC	1380
Dd	1321	CCCGCAAGAA	GACACAGAAAGAGCCCCCTCTCTGTGATGAGCTGCACCCC	1380
Qy	1381	GACAAATG	AGACCCGTGACGCCCTTGAGACTGCCGAGAAAGAGAGCTGGAACGAC	1440
Dd	1381	GACAAATG	AGACCCGTGACGCCCTTGAGACTGCCGAGAAAGAGAGCTGGAACGAC	1440
Qy	1441	ATCCGAGA	AGCTGAGGCGCAAGCTGAACTGAGGCGCAGACCTAACCCCGCATCAAGGTG	1500
Dd	1441	ATCCGAGA	AGCTGAGGCGCAAGCTGAACTGAGGCGCAGACCTAACCCCGCATCAAGGTG	1500
Qy	1501	CGCAGCTG	TGACAGCTGTCGCGCGCGCCAGAGCCCTGACCGACATCGTACCCTTGACC	1560
Dd	1501	CGCAGCTG	TGACAGCTGTCGCGCGCGCCAGAGCCCTGACCGACATCGTACCCTTGACC	1560
Qy	1561	GAGAGAGC	CCGAGCTGAGAGCTGAGCTGAGCTGAGAACCGCGAGATCTCTGCGCAGAGCCCTGTCAGCGC	1620
Dd	1561	GAGAGAGC	CCGAGCTGAGAGCTGAGCTGAGCTGAGAACCGCGAGATCTCTGCGCAGAGCCCTGTCAGCGC	1620
Qy	1621	GTCATCTA	ACGACCCCAACAAAGAACTGTGTGGCGAGATCCAGAAAGCAGAGGCGACAGCAG	1680
Dd	1621	GTCATCTA	ACGACCCCAACAAAGAACTGTGTGGCGAGATCCAGAAAGCAGAGGCGACAGCAG	1680
Qy	1681	TGACCTTA	CCAGATCTACAGAGAGCCCTTCAAGAACCTTGAAAGACCCGACAGTACCGCAAG	1740
Dd	1681	TGACCTTA	CCAGATCTACAGAGAGCCCTTCAAGAACCTTGAAAGACCCGACAGTACCGCAAG	1740
Qy	1741	ATGCGCAC	CCGCGCACACCAACAGATGTAGACGTGACCGAGGCGCTGTGCAGAAAGATGCGC	1800
Dd	1741	ATGCGCAC	CCGCGCACACCAACAGATGTAGACGTGACCGAGGCGCTGTGTGAGAAAGATGCGC	1800
Qy	1801	ATGAGAGCA	TCGTGATCTGTGGGCAAAACCCCAAGTTCCGCGCTGCGCCATCCAGAGAGAG	1860
Dd	1801	ATGAGAGCA	TCGTGATCTGTGGGCAAAACCCCAAGTTCCGCGCTGCGCCATCCAGAGAGAG	1860
Qy	1861	ACCTGAGAG	ACCTGAGAGACCTGAGTGAACCTGAGACCGACCTTGATCCCGAGTGGAGATTTC	1920
Dd	1861	ACCTGAGAG	ACCTGAGAGACCTGAGTGAACCTGAGACCGACCTTGATCCCGAGTGGAGATTTC	1920
Qy	1921	GTGAAAC	CCCCCTGTGTGAGACTGTGTGTACAGCTGTGAGAAAGAGGCCCATCTGCGC	1980
Dd	1921	GTGAAAC	CCCCCTGTGTGAGACTGTGTGTACAGCTGTGAGAAAGAGGCCCATCTGCGC	1980
Qy	1981	GCGAGACT	TTTCTACGTGTGAGACCGCGCGCGCAACCGCGAGACCAAGATCCGACAGAGCGCGC	2040
Dd	1981	GCGAGACT	TTTCTACGTGTGAGACCGCGCGCGCAACCGCGAGACCAAGATCCGCGAGAGCGCGC	2040
Qy	2041	TACGTGA	CCCGACCGGCGCGCGCAGAGATCGTGAAGCTGACCCGAGACCAACCAAG	2100
Dd	2041	TACGTGA	CCCGACCGGCGCGCGCAGAGATCGTGAAGCTGACCCGAGACCAACCAAG	2100
Qy	2101	ACCGAGCT	GACGCGCATTCAGACTGCGCTGTGAGGACAGCGGACGAGGTGAACATCTGTG	2160
Dd	2101	ACCGAGCT	GACGCGCATTCAGACTGCGCTGTGAGGACAGCGGACGAGGTGAACATCTGTG	2160
Qy	2161	ACCGACAG	CCGACGCTGCGCGCTGTGAGGACATCAAGGCGCCGACCAAGAGCGAGAGGAG	2220
Dd	2161	ACCGACAG	CCGACGCTGCGCGCTGTGAGGACATCAAGGCGCCGACCAAGAGCGAGAGGAG	2220
Qy	2221	CTGTGTGA	CCAGATCTGAGCAGCTGATCAAGAAAGAGAGAGGTGACTGTGAGCTGTGGTG	2280
Dd	2221	CTGTGTGA	CCAGATCTGAGCAGCTGATCAAGAAAGAGAGAGGTGACTGTGAGCTGTGGTG	2280
Qy	2281	CCGCGCCA	CAAGGCGCATCGCGCGCAACGAGCGATGCACAAGCTGTGAGCAAGGCGATC	2340
Dd	2281	CCGCGCCA	CAAGGCGCATCGCGCGCAACGAGCGATGCACAAGCTGTGAGCAAGGCGATC	2340
Qy	2341	CGCAAGTG	CTGTTCTGTGACGCGCATGTAGTGCGCGCATCTGTATCTACAGATCACTGAC	2400
Dd	2341	CGCAAGTG	CTGTTCTGTGACGCGCATGTAGTGCGCGCATCTGTATCTACAGATCACTGAC	2400

QY 1621 GTGTACTAGACCCGAGCAGAGGACTGTGTGCGGAGATTCAGAGACAGGCGCAGCAG 1680
 DB 1615 GTGTACTAGACCCGAGCAGAGGACTGTGTGCGGAGATTCAGAGACAGGCGCAGCAG 1674
 QY 1661 TGGACTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACCCAG 1740
 DB 1675 TGGACTTACAGATCTTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACCCAG 1734
 QY 1741 ATGCGCACCGGCGCAGCAGCAGAGCTGAGAGCAGCTGACCGAGCCGTGCAAGATCCGC 1800
 DB 1735 ATGCGCACCGGCGCAGCAGCAGAGCTGAGAGCAGCTGACCGAGCCGTGCAAGATCCGC 1794
 QY 1801 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAG 1860
 DB 1795 ATGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAG 1854
 QY 1861 ACCTGGAGAGACTGTGTGACCGCATCTAGGAGGCCACTTGATCCCGAGTGGAGTTTC 1920
 DB 1855 ACCTGGAGAGACTGTGTGACCGCATCTAGGAGGCCACTTGATCCCGAGTGGAGTTTC 1914
 QY 1921 GTGAACACCCCGCTGTGTGAGCTGTGTACAGCTGAGAGAGAGCCCATCCGAC 1980
 DB 1915 GTGAACACCCCGCTGTGTGAGCTGTGTGTACAGCTGAGAGAGAGCCCATCCGAC 1974
 QY 1981 GCGGAGACCTTCTACGTGTGACCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCCGAC 2040
 DB 1975 GCGGAGACCTTCTACGTGTGACCGCGCGCCCAACCGGAGACCAAGATCGGCAAGCCGAC 2034
 QY 2041 TACGTGACCGACCGGCG 2100
 DB 2035 TACGTGACCGACCGGCG 2094
 QY 2101 ACCGAGCTGACAGGCGCATCCAGCTGCGCTGTGACGACAGCGGCGAGCGGCTGAACTCGTG 2160
 DB 2095 ACCGAGCTGACAGGCGCATCCAGCTGCGCTGTGACGACAGCGGCGAGCGGCTGAACTCGTG 2154
 QY 2161 ACCGAGCAGCATGACGCTGTGGCATCATCCAGCGCCAGCCCGCAAGAGCGAGCGAG 2220
 DB 2155 ACCGAGCAGCATGACGCTGTGGCATCATCCAGCGCCAGCCCGCAAGAGCGAGCGAG 2214
 QY 2221 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGTG 2280
 DB 2215 CTGTGTGAACCAAGATCATCGAGCAGCTGATCAAGAGAGAGAGTGTACTTGAAGCTGGTG 2274
 QY 2281 CCGCGCCAGAGGCGCATCGGCGGCGCAACAGCAGATCGACAGCTGGTGAAGAGGCGCATC 2340
 DB 2275 CCGCGCCAGAGGCGCATCGGCGGCGCAACAGCAGATCGACAGCTGGTGAAGAGGCGCATC 2334
 QY 2341 CGCAAGGTGTCTTCTGTGACCGGCATCGATGAGCGGCGCATGTGATCTACCAAGTACATGAGC 2400
 DB 2335 CGCAAGGTGTCTTCTGTGACCGGCATCGATGAGCGGCGCATGTGATCTACCAAGTACATGAGC 2394
 QY 2401 GACCTGTACGTGGGCGAGCGGCGGCGCTTGAAGATCGATTAAAGCTTCCGGGCGTACGACC 2460
 DB 2395 GACCTGTACGTGGGCGAGCGGCGGCGCTTGAAGATCGATTAAAGCTTCCGGGCGTACGACC 2454
 QY 2461 GGTGAATTC 2469
 DB 2455 GGTGAATTC 2463

RESULT 3
 AX455916 2457 bp DNA linear PAT 06-JUL-2002
 LOCUS AX455916
 DEFINITION Sequence 32 from Patent WO0204493.
 ACCESSION AX455916
 VERSION AX455916.1 GI:21714901
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct
 other sequences; artificial sequences.
 REFERENCE
 1 zur Megeide, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
 JOURNAL polyptides and uses thereof
 CHIRON CORPORATION (US) ; Universality of Stellenbosch (ZA)
 FEATURES
 source
 1. 2457
 /organism="synthetic construct"
 /mol type="unassigned DNA"
 /db xref="cazon:32630"
 /note="PR975YMMM"
 ORIGIN
 Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
 Best Local Similarity 99.3%; Pred. No. 2.7e-199;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
 QY 1 GTGAGCGCACCATGTGGCGAGGCGCATGTAGGCCACAGGCGCAACATCTGTATGCG 60
 DB 1 GTGAGCGCACCATGTGGCGAGGCGCATGTAGGCGCACAGCGCGCAACATCTGTATGCG 60
 QY 61 CGCAGCACTTCAAGAGGCGCCCAAGGCGCATCATAGTGTCTCAACTGCGGCGAAGAGGCGC 120
 DB 61 CGCAGCACTTCAAGAGGCGCCCAAGGCGCATCATAGTGTCTCAACTGCGGCGAAGAGGCGC 120
 QY 121 CACATGCGCGCACTGCG 180
 DB 121 CACATGCGCGCACTGCG 180
 QY 181 GCGCACAGATGAGAGACTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
 DB 181 GCGCACAGATGAGAGAGACTGTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
 QY 241 TTCCCCCAGAGGCGAAGCG 300
 DB 241 TTCCCCCAGAGGCGAAGCG 300
 QY 301 AGCCGAGCTGTGACAGTGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCG 360
 DB 301 AGCCGAGCTGTGACAGTGTGCGCGCGCAACCCCGCGAGCGAGCGCGCGCGCGCGCGCG 360
 QY 361 GGCACCTGTAACTTCCCCAGATCAACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
 DB 361 GGCACCTGTAACTTCCCCAGATCAACCTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
 QY 421 GCGCGCGAGATCAAGAGGCG 480
 DB 421 GCGCGCGAGATCAAGAGGCG 480
 QY 481 ATGAGCTGTGCGCGAGGTGAGAGCGCGCAAGATGATCGGCGCGCATCGGCGGCTTCAATCAAG 540
 DB 481 ATGAGCTGTGCGCGAGGTGAGAGCGCGCAAGATGATCGGCGCGCATCGGCGGCTTCAATCAAG 540
 QY 541 GTGCGCGAGTACCAAGATCTGTATCGAGATCTGTGCGCGCGCGCGCGCGCGCGCGCGCG 600
 DB 541 GTGCGCGAGTACCAAGATCTGTATCGAGATCTGTGCGCGCGCGCGCGCGCGCGCGCGCG 600
 QY 601 CTGATGCG 660
 DB 601 CTGATGCG 660
 QY 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGGTGCGCGCGCGCGCGCGCGCGCGCG 720
 DB 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGGTGCGCGCGCGCGCGCGCGCGCGCGCG 720
 QY 721 GAGCG 780
 DB 721 GAGCG 780
 QY 781 ATCTGCGAGAGATGAGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
 DB 781 ATCTGCGAGAGATGAGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
 QY 841 AACACCCCGTGTGCGCATCAAG 900

Db 841 AAGACCCCGTGTGGCCATCAAGAAAGAGACCAAGTGGCGCAAGCTGGAGC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCAACCAAGACTTCTGGAGAGTGCAGCTGGGATCCCGAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCAACCAAGACTTCTGGAGAGTGCAGCTGGGATCCCGAC 960
Qy 961 CCCGCGGCTTGAAGAAAGAGAGCTGACCTGTGCTGAAGTGGGGCGACCTTACTTC 1020
Db 961 CCCGCGGCTTGAAGAAAGAGAGCTGACCTGTGCTGAAGTGGGGCGACCTTACTTC 1020
Qy 1021 AGCGTGGCCCTTGAAGAGAGCTTCCGCAAGTACACCGGCTTACCAATCCCGACATCAAC 1080
Db 1021 AGCGTGGCCCTTGAAGAGAGCTTCCGCAAGTACACCGGCTTACCAATCCCGACATCAAC 1080
Qy 1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAGTGTGCTCCCAAGGCTGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAGTGTGCTCCCAAGGCTGAAGGGCAGC 1140
Qy 1141 CCCAGATTTTCAAGAGCAATGACCAAGATCTGAGAGCTTCCGCGCGCGCAACCC 1200
Db 1141 CCCAGATTTTCAAGAGCAATGACCAAGATCTGAGAGCTTCCGCGCGCGCAACCC 1200
Qy 1201 GAGATCTGTATCTACCA-----GGCGCCCTGTATGAGGCGACCTGAGATCGGC 1260
Db 1201 GAGATCTGTATCTACCA-----GGCGCCCTGTATGAGGCGACCTGAGATCGGC 1260
Qy 1261 CAGCACCGCGCAAGATCGAGAGAGTGCAGCAAGCACTGTGCGTGGGGCTTACCAAC 1320
Db 1255 CAGCACCGCGCAAGATCGAGAGAGTGCAGCAAGCACTGTGCGTGGGGCTTACCAAC 1314
Qy 1321 CCCGACAGAGCAACCAAGAGAGCGCCCTTCTGTGATGGGCTTACGAGCTGCACCC 1380
Db 1315 CCCGACAGAGCAACCAAGAGAGCGCCCTTCTGTGATGGGCTTACGAGCTGCACCC 1368
Qy 1381 GACAGTGGAGCGGTGAGCGCCATCGAGCTGCCCGAAGAGAGAGTGGACCGTGAAGCAG 1440
Db 1369 GACAGTGGAGCGGTGAGCGCCATCGAGCTGCCCGAAGAGAGAGTGGACCGTGAAGCAG 1428
Qy 1441 ATTCAGAAAGCTGTGGGCAAGCTGAAGTGGGCGCAAGCACTGATCCCGGCAATCAAGTG 1500
Db 1429 ATTCAGAAAGCTGTGGGCAAGCTGAAGTGGGCGCAAGCACTGATCCCGGCAATCAAGTG 1488
Qy 1501 CGCGAGCTGTGCAAGCTGTGCGCGCGCGCAAGAGCCCTGAACCGCAATGTGCCCTGACC 1560
Db 1489 CGCGAGCTGTGCAAGCTGTGCGCGCGCGCGCAAGAGCCCTGAACCATGTGCCCTGACC 1548
Qy 1561 GAGAGAGCGGAGCTGAGAGCTGGCGGAGAACCGGAGATCTGTGGCGAGCCCGTGCACGCG 1620
Db 1549 GAGAGAGCGGAGCTGAGAGCTGGCGGAGAACCGGAGATCTGTGGCGAGCCCGTGCACGCG 1608
Qy 1621 GTGTACTAGCAAGCCCGCAAGAGACTGTGGCGGAGATCCAGAGAGAGGCGCACAGCAG 1680
Db 1609 GTGTACTAGCAAGCCCGCAAGAGACTGTGGCGGAGATCCAGAGAGAGGCGCACAGCAG 1668
Qy 1681 TGGACCTTACAGATCTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAG 1740
Db 1669 TGGACCTTACAGATCTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAG 1728
Qy 1741 ATGGCGACCGGCGCAACCAAGAGCTGGAAGAGCTGACCGAGGCGCTGCAAGAAATCGGC 1800
Db 1729 ATGGCGACCGGCGCAACCAAGAGCTGGAAGAGCTGACCGAGGCGCTGCAAGAAATCGGC 1788
Qy 1801 ATGAGAGAGCTGTGATCTGGGGCAAGAGCCCGCAAGTCCGCTGCGCTTCCAGAGAGAG 1860
Db 1789 ATGAGAGAGCTGTGATCTGGGGCAAGAGCCCGCAAGTCCGCTGCGCTTCCAGAGAGAG 1848
Qy 1861 ACCTGGAGAGCTGTGAGAGAGCTACTGTGAGAGGCGCACTGTATCCCGAGTGGAGTTTC 1920
Db 1849 ACCTGGAGAGCTGTGAGAGAGCTACTGTGAGAGGCGCACTGTATCCCGAGTGGAGTTTC 1908
Qy 1921 GTGAGAACCCCGCGCTGTGAGAGCTGTGTACTAGCTGAGAGAGAGCCCATCTCGCG 1980

Db 1909 GTGAGAACCCCGCGCTGTGAGAGCTGTGTACTAGCTGAGAGAGAGCCCATCTCGCG 1968
Qy 1961 GCCGAGACCTTCTTACTGTGAGAGCGCGCGCGCAACCGCGAGAGCTTAAGATGGGCAAGCGCGCG 2040
Db 1969 GCCGAGACCTTCTTACTGTGAGAGCGCGCGCGCGCAACCGCGAGAGCTTAAGATGGGCAAGCGCGCG 2028
Qy 2041 TAAGTGAAGAGAGCGGCGCGCGAGAGAGTGTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db 2029 TAAGTGAAGAGAGCGGCGCGCGAGAGAGTGTGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
Qy 2101 ACCGAGCTGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db 2089 ACCGAGCTGAGAGAGAGAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2148
Qy 2161 ACCGAG 2220
Db 2149 ACCGAG 2208
Qy 2221 CTGTGAG 2280
Db 2209 CTGTGAG 2268
Qy 2281 CCCGCGCAAG 2340
Db 2269 CCCGCGCAAG 2328
Qy 2341 CGCAAGTGTGTCTTGTGAG 2400
Db 2329 CGCAAGTGTGTCTTGTGAG 2388
Qy 2401 GACCTGTAGTGTGGAG 2460
Db 2389 GACCTGTAGTGTGGAG 2448
Qy 2461 GGTGAATTC 2469
Db 2449 GGTGAATTC 2457

RESULT 4
BD263706 2312 bp DNA linear PART 17-JUL-2003
LOCUS
DEFINITION Improved expression of HIV polypeptides and production of
virus-like particles.
ACCESSION BD263706
VERSION BD263706.1 GI:33073474
KEYWORDS JP 2002533124-A/73.
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 2312)
Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
Improved expression of HIV polypeptides and production of
virus-like particles
Patent: JP 2002533124-A 73 08-OCT-2002;
TITLE
JOURNAL
CHIRON CORP
COMMENT
OS Artificial Sequence
PN JP 2002533124-A/73
PD 08-OCT-2002
PP 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIWASTAVA, YING LIAN, PI
KARIN HARTOG,
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protomod.Rtopr(+) FH
Location/Qualifiers
FT source 1..2312
FT Location/Qualifiers /organism='Artificial Sequence'.

SOURCE

1. 2312
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"

ORIGIN

```
Query Match      83.1%; Score 2052; DB 6; Length 2312;
Beet Local Similarity 93.6%; Pred. No. 4,9e-168;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

170 GCGGCAAGAGGGCCCAAGATGAGAGTGCACCGAGGCGCAAGCTTCTCCGG 229
    1 GCGGCGCGAAGAGCAACCAATGAAAGATTGCACTGAGAGCAGAGCTTAATTTCTCCGG 60
    230 AGGACTTGAGCTTCTTCGAGGCGCAAGGCCCGGAGATTCCCGAGCGAGCAACCGGCCA 289
    61 AGGACTTGAGCTTCTTCGAGGCGCAAGGCCCGGAGATTCCCGAGCGAGCAACCGGCCA 120
    230 AAGAGCCCAACAGCCCGAGCTGCAAGGTGCGGCG-----ACAACCCCGAGCGAGG 343
    121 AAGAGCCCAACAGCCCGAGCTGCAAGGTGCGGCGCGAGAAACAAGCTGAGCGAGG 180
    344 CCGGCGCGAGGCGCGAGGCAACCTG-----AATTCCCGCAATCACTCTGTGCAAG 397
    181 CCGGCGCGAGGCGCGAGGCAACCTGCAACTTCCCGCAATCACTCTGTGCAAG 240
    398 GCCCCTGTGAGCATCAAGGTGCGGCGCGAGATCAAGAGGCCCTGTGCAACCGGCG 457
    241 GCCCCTGTGAGCATCAAGGTGCGGCGCGAGATCAAGAGGCCCTGTGCAACCGGCG 300
    458 CCGAGCAACCGGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 517
    301 CCGAGCAACCGGTGTGAGAGAGATGAACTGCGCGCAAGTGAAGCCCAAGATGATCG 360
    518 GCGGCAATGCGCGCTTTCATCAAGGTGCGCAAGTGAAGCAAGTCTGATGAGATCTGCG 577
    351 GCGGCAATGCGCGCTTTCATCAAGGTGCGCAAGTGAAGCAAGTCTGATGAGATCTGCG 420
    578 GCAAGAGGCGCATCGGCAACGTGTGATGCGCCCAACCCCGTGAACATCATCGGCGCA 637
    421 GCCAAGAGCGCATCGGCAACGTGTGATGCGCCCAACCCCGTGAACATCATCGGCGCA 480
    638 AATGCTGACCCAGTGGGCTGCAACCTGAACTTCCCATCAAGCCCATGAGACCGTGC 697
    481 AATGCTGACCCAGTGGGCTGCAACCTGAACTTCCCATCAAGCCCATGAGACCGTGC 540
    698 CCGTGAAGCTGAAGCCCGGATGAGACGCGCCCAAGTGAAGCAAGTGGCCCTTGAACGAG 757
    541 CCGTGAAGCTGAAGCCCGGATGAGACGCGCCCAAGTGAAGCAAGTGGCCCTTGAACGAG 600
    758 AGAAGATCAAGGCCCTGACCGGCATCTGCAAGAGATGAGAAAGGCGCAAGATCAACA 817
    601 AGAAGATCAAGGCCCTGAGTGAAGATCTGCAAGAGTGAAGAGGCGCAAGATCAACA 660
    818 AGATCGGCCCGGAGAACCCCTTCAACAACCCCGTGTTCGCATCAAGAAAGAGACAGA 877
    661 AGATCGGCCCGGAGAACCCCTTCAACAACCCCGTGTTCGCATCAAGAAAGAGACAGA 720
    878 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGCTGAACAAGCGCAACCCAGACTTCTGCG 937
    721 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGCTGAACAAGCGCAACCCAGACTTCTGCG 780
    938 AGGTGAGCTGAGGATCCCGCAACCCCGCGGCGTGAAGAAAGAAAGACCGTGAACCGTGC 997
    781 AGGTGAGCTGAGGATCCCGCAACCCCGCGGCGTGAAGAAAGAAAGACCGTGAACCGTGC 840
    998 TGAAGTGGCGCAAGCTGTGAGCTTCAAGCGTGCCTGTGACGAGACTTCCGCAAGTCAACCG 1057
    841 TGAAGTGGCGCAAGCTGTGAGCTTCAAGCGTGCCTGTGACGAGACTTCCGCAAGTCAACCG 900
    1058 CTTTCAACCATCCCGAGATGAACAACAGAGACCCCGGCAATCCGCTACCAAGTGC 1117
    901 CTTTCAACCATCCCGAGATGAACAACAGAGACCCCGGCAATCCGCTACCAAGTGC 960
```

```
1118 TGCCCCAGAGGCTGAAAGGAGAGCCCGAGCATCTTCCAGAGCAGATGACCAAGATCTTGG 1177
    961 TGCCCCAGAGGCTGAAAGGAGAGCCCGGAGCATCTTCCAGAGCAGATGACCAAGATCTTGG 1020
    1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCTGATCTCAAGTCAATGACGACCTGTAGC 1237
    1021 AGCCCTTCCGCGAGAGAACCCCGACATCTGATCTCAAGTCAATGACGACCTGTAGC 1080
    1238 TGAGGAGGACCTTGAGATTCGGCGAGCAACCGGCGCAAGATGAGAGAGCTGCGCAAGC 1297
    1081 TGAGGAGGACCTTGAGATTCGGCGAGCAACCGGCGCAAGATGAGAGAGCTGCGCGAGC 1140
    1298 TGCTGCGTGGGAGCTTCAACAACCCCGCAAGAAACAACAGAGAGCCCTTCTGT 1357
    1141 TGCTGCGTGGGAGCTTCAACAACCCCGCAAGAAACAACAGAGAGCCCTTCTGT 1200
    1358 GATGAGGCTACGAGCTGCAACCCCGCAAGTGAACCTGTGAGCCATCGAGCTGCGGAGA 1417
    1201 GATGAGGCTACGAGCTGCAACCCCGCAAGTGAACCTGTGAGCCATCGAGCTGCGGAGA 1260
    1418 AGGAGAGCTGACCTGTGAACCAATCCAGAAAGTGTGGGCAAGCTGAACTGGGCGAGCC 1477
    1261 AGGAGAGCTGACCTGTGAACCAATCCAGAAAGTGTGGGCAAGCTGAACTGGGCGAGCC 1320
    1478 AGATCTAACCCCGGCAATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCAAGGCC 1537
    1321 AGATCTAACCCCGGCAATCAAGGTGCGCGAGCTGTGCAAGCTGTGCGCGCGCAAGGCC 1380
    1538 TGAACGAGATGCTGCGCTTGAACGAGAGGCCGAGCTGAGCTGAGCCGAGAACCGCGAGA 1597
    1381 TGAACGAGATGCTGCGCTTGAACGAGAGGCCGAGCTGAGCTGAGCCGAGAACCGCGAGA 1440
    1598 TCTGCGGAGAGCTGCGCTGCAACGCGGTGTAATCAAGACCCCGAGAAAGCACTGTGGCGAGA 1657
    1441 TCTGCGGAGAGCTGCGCTGCAACGCGGTGTAATCAAGACCCCGAGAAAGCACTGTGGCGAGA 1500
    1658 TCCAGAGAGGAGCGCAACGAGCTGAGCTCAAGATCTTACAGAGAGCCCTTCAAGAAC 1717
    1501 TCCAGAGAGGAGCGCAACGAGCTGAGCTCAAGATCTTACAGAGAGCCCTTCAAGAAC 1560
    1718 TGAAGACCGGAGAGTATGCGCAAGATGAGCAACCGCCCAACAACGAGAGAGCTGTA 1777
    1561 TGAAGACCGGAGAGTATGCGCAAGATGAGCAACCGCCCAACAACGAGAGAGCTGTA 1620
    1778 CCGAGGCGGTGCAAGAGATGCGCATGAGAGAGATGTGTATCTGGGCGAAGACCCCAAGT 1837
    1621 CCGAGGCGGTGCAAGAGATGAGCAACGAGAGATGTGTATCTGGGCGAAGATCCCAAGT 1680
    1838 TCCGCTGCGCATTCAGAAAGAGACCTGAGAGACCTGTGAGACGATCTGAGAGAGCA 1897
    1681 TCAAGCTGCGCATTCAGAAAGAGACCTGAGAGAGCTGTGAGATGAGAGAGCA 1740
    1898 CTTGATCCCGAGTGGAGATTCTGTAACAACCCCGCTGTGTGAAGCTGTGTACAGC 1957
    1741 CTTGATCCCGAGTGGAGATTCTGTAACAACCCCGCTGTGTGAAGCTGTGTACAGC 1800
    1958 TGAAGAGAGGCTCATCATGCGCGCGAGACCTTCTAGTGAACGCGCGCGCAACCGCG 2017
    1801 TGAAGAGAGGCTCATCATGCGCGCGAGACCTTCTAGTGAACGCGCGCGCAACCGCG 1860
    2018 AGAACAAGATGGGCAAGGCGGCTTACGTGAACGAGCCGAGGCGGAGAAAGATCTGAGGC 2077
    1861 AGAACAAGATGGGCAAGGCGGCTTACGTGAACGAGCCGAGGCGGAGAAAGATCTGAGGC 1920
    2078 TGAACGAGACCAACAACAGAGACCGAGCTGCAAGCCATCCAGCTGCGCTTCAAGACA 2137
    1921 TGAACGAGACCAACAACAGAGACCGAGCTGCAAGCCATCCAGCTGCGCTTCAAGACA 1980
    2138 GCGGCGAGAGGTGAACATCTGAGACGACCAATGAGCTTGGGCAATCTCAAGGCC 2197
    1981 GCGGCGAGAGGTGAACATCTGAGACGACCAATGAGCTTGGGCAATCTCAAGGCC 2040
```



```
QY 1598 TCCTGCGAGGCGCGTGCACGCGGTACTACGACCCCGACGAAGGACCTGTGGCGGAGA 1657
DB 1441 TCCTGAGAGAGCGCTGTGCACGAGGTACTACGACCCCGACGAAGGACCTGTGGCGGAGA 1500
QY 1658 TCAGAGAGAGGCGGACGACGAGTGTGACCTTACGAGATCTTACGAGAGCGCTTCAAGAAC 1717
DB 1501 TCAGAGAGAGGCGGACGAGGCGCAGTGTGACCTTACGAGATCTTACGAGAGCGCTTCAAGAAC 1560
QY 1718 TGAAGACCGGAGAGTACGCGAAGATGTGACCGCGCCACGACGACGAGTGTGAGAGGAGCTGA 1777
DB 1561 TGAAGACCGGAGAGTACGCGCGATGTGACGCGCGCCACGACGAGTGTGAGAGGAGCTGA 1620
QY 1778 CCGAGGCGGTGACGAAGATCGCGCATGTGAGAGCATGTGTCTGTGGGCAAGACCCCAAGT 1837
DB 1651 CCGAGGCGGTGACGAAGGTGAGACCGAGAGGATGTGTATCTGTGGGCAAGATCCCAAGT 1680
QY 1838 TCCTGCTGCGCATTCAGAGAGAGACCTGTGAGACCTGTGTGAGCCGACTTACGAGGCGCA 1897
DB 1681 TCAGAGCTGCGCATTCAGAGAGAGACCTGTGAGAGCGCTGTGTGAGATGTGTGAGGCGCA 1740
QY 1898 CCGTGTATCCCGAGTGTGGAGTCTGTGAGACACCCCGCGCTGTGTGAGAGCTGTGTGAGC 1957
DB 1741 CCGTGTATCCCGAGTGTGGAGTCTGTGAGACACCCCGCGCTGTGTGAGAGCTGTGTGAGC 1800
QY 1958 TGAAGAGAGAGCGCCATCATCTGCGCGCGGACCTTCTACGTGTGACGCGCGCGCAACCGCG 2017
DB 1801 TGAAGAGAGAGCGCCATCATCTGCGCGCGGACCTTCTACGTGTGACGCGCGCGCGCAACCGCG 1860
QY 2018 AGACCAAGATTCGCGCAAGGCGCGCTTACGTGACCGACCGCGCGCGCGCAAGAGATGTGAGCG 2077
DB 1861 AGACCAAGATTCGCGCGCAAGGCGCGCTTACGTGACCGACCGCGCGCGCGCAAGAGATGTGAGCG 1920
QY 2078 TGAACGAGACCGACCGACCGAGAGACCGAGCTGTGAGAGCGCCATTCGAGCTGTGCGCTGTGAGACA 2137
DB 1921 TCGCGGACCGACCGACCGAGAGACCGAGCTGTGAGAGCGCCATTCGAGCTGTGCGCGCTGTGAGACA 1980
QY 2138 GCGGACGAGCGAGTGAACATCTGTGACCGACCGAGCGCATTCGAGCGCGCGCGCGCGCGCG 2197
DB 1981 GCGGACGAGCGAGTGAACATCTGTGACCGACCGAGCGCATTCGAGCGCGCGCGCGCGCGCGCG 2040
QY 2198 AGCCCGACCAAGAGCGAGAGCGAGCTGTGAGACCGAGCTGTGAGAGCGCTGTGAGAGAGG 2257
DB 2041 AGCCCGACCAAGAGCGAGAGCGAGCTGTGAGACCGAGCTGTGAGAGCGCTGTGAGAGAGG 2100
QY 2258 AGAAGGTGTACTGTGAGCTGTGGTGTGCGCCCGACCAAGGCGCATTCGCGCGCGCGCGCGCGCG 2317
DB 2101 AGAAGGTGTACTGTGAGCTGTGGTGTGCGCCCGACCAAGGCGCATTCGCGCGCGCGCGCGCGCG 2160
QY 2318 ACAAGCTGTGAGCAAGGCGCATTCGCGAGAGTGTGTCTGTGAGCGCGCATTCGAGCGCGCA 2377
DB 2161 ACAAGCTGTGAGCAAGGCGCGCGCGCATTCGCGAGAGTGTGTCTGTGAGCGCGCATTCGAGCGCGCA 2220
QY 2378 TCGTGTACTTACGAGTACATGTGACGACCTGTGAGTGTGGCGCGCGCGCGCGCGCGCGCGCGT 2437
DB 2221 TCGTGTACTTACGAGTACATGTGACGACCTGTGAGTGTGGCGCGCGCGCGCGCGCGCGCGCGT 2280
QY 2438 AAAAGCTTCCCGGCGGTGAGACCGCGGTAATTC 2469
DB 2281 AAAAGCTTCCCGGCGGTGAGACCGCGGTAATTC 2312
```

```
RESULT 6
LOCUS AR373389 2312 bp DNA linear PART 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
```

```
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source
location/Qualifiers
1..2312
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 83.1%; Score 2052; DB 6; Length 2312;
Best Local Similarity 93.6%; Pred. No. 4.9e-168;
Matches 2165; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
QY 170 GCGGAGAGAGGCGGACCGAGTGTGAGAGCTGTGACCGGAGCGCGCAAGCGCGCACTTCTTCCCG 229
DB 1 GCGGCGCGGAGAGAGACCAAGATGTGACCTGTGAGAGCGGCTAATTTCTTCCCG 60
QY 230 AGGACCTGCGCTTCCCGAGGCGCAAGGCGCGGAGTTCGCCAGCGAGAGAACCGCGCA 289
DB 61 AGGACCTGCGCTTCTGTGAGGCGCAAGGCGCGGAGTTCAGAGCGAGAGAGCGCGCGCA 120
QY 290 ACAAGCCCAACGAGCGGAGCTGTGAGTGTGCGCGCG-----ACAAGCCCGAGCGAGG 343
DB 121 ACAAGCCCAACGAGCGGAGCTGTGAGTGTGCGCGCGAGAGAGCAAGAGCTGTGAGAGG 180
QY 344 CCGGCGCGGAGCGGAGGAGCGCGCG-----AATTTCGCCAGTGTGAGTGTGAGCG 397
DB 181 CCGGCGCGGAGCGGAGGAGCGCGCGAGCTTCACTTCCCGAGTGTGAGTGTGAGCG 240
QY 398 GCCCGCTGTGAGATCAAGGTGTGCGCGCGAGATGTGAGAGGCGCTGTGTGAGCAACCGCG 457
DB 241 GCCCGCTGTGAGATCAAGGTGTGCGCGCGAGCTGTGAGAGGCGCTGTGTGAGCAACCGCG 300
QY 458 CCGAGAGACCGGTGTGAGAGAGATGTGAGCTGTGCGCGAGATGTGAGAGCGCGAGATGTG 517
DB 301 CCGAGAGACCGGTGTGAGAGAGATGTGAGCTGTGCGCGAGATGTGAGAGCGCGAGATGTG 360
QY 518 GCGGATGTGCGGCTTCACTCAAGGTGTGCGCGCGAGATGTGAGAGTGTGAGTGTGAG 577
DB 351 GCGGATGTGCGGCTTCACTCAAGGTGTGCGCGCGAGATGTGAGAGTGTGAGTGTGAG 420
QY 578 GCAAGAGAGCGCATTCGCGCGCGGTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 637
DB 421 GCGAGAGAGCGCATTCGCGCGCGGTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 638 ACATGTGTGACCGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 697
DB 481 ACATGTGTGACCGAGTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 540
QY 698 CCGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
DB 541 CCGTGTGAGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
QY 758 AGAAGATCAAGGCGCTGTGAGCGCGCATTCGAGAGAGTGTGAGAGAGAGAGAGAGAGTCA 817
DB 601 AGAAGATCAAGGCGCTGTGAGCGCGCATTCGAGAGAGTGTGAGAGAGAGAGAGAGTCA 660
QY 818 AGATGTGCGCGCGGAGAGCGCGTGTGAGACACCGCGGTGTGCGCATCAAGAGAGAGAGCA 877
DB 661 AGATGTGCGCGCGGAGAGCGCGTGTGAGACACCGCGGTGTGCGCATCAAGAGAGAGAGCA 720
QY 878 CCAAGTGTGCGGAGCTGTGAGCTTCCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGTCT 937
DB 721 CCAAGTGTGCGGAGCTGTGAGCTTCCGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGTCT 780
QY 938 AGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 997
DB 781 AGGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 840
QY 998 TGAAGTGTGAGCGCGCTTCACTTCAAGCTGTGCGCGTGTGAGAGAGAGTGTGAGAGTCA 1057
```

Db 841 TGGACGTGGGCGACGCTTACTTACGGTGCCTTGGACAAAGACTTCCGCAAGTACACCG 900
 QY 1058 CCTTCCATCCCGACGATCAACACAGAACCCCGGCATCCGCTTACAGTACACGCTGC 1117
 Db 901 CCTTCCATCCCGACGATCAACACAGAACCCCGGCATCCGCTTACAGTACACGCTGC 960
 QY 1118 TGGCCCGAGGCTGGAGAGGCGCGCGAGCTTTCCAGAGCAGTACCAAGATCTCTG 1177
 Db 961 TGGCCCGAGGCTGGAGAGGCGCGCGAGCTTTCCAGAGCAGTACCAAGATCTCTG 1020
 QY 1178 AGCCCTTCCGCGCCGCGCAACCCCGAGTGTATCTACCAAGTACATGAGCAAGCTTGTAG 1237
 Db 1021 AGCCCTTCCGCGCAAGCAAGAACCCCGCATGTATCTACCAAGTACATGAGCAAGCTTGTAG 1080
 QY 1238 TGGGACGCACTTGAAGATCGGCGACACCGCGCAAGATCGAGAGCTTGGCAAGCACCC 1297
 Db 1081 TGGGACGCACTTGAAGATCGGCGACACCGCGCAAGATCGAGAGCTTGGCGACACCC 1140
 QY 1298 TGTGTGCTGGGGCTTACCAACCCCGCAAGAGCAACCAAGAGAGCCCTTCTGT 1357
 Db 1141 TGTGTGCTGGGGCTTACCAACCCCGCAAGAGCAACCAAGAGAGCCCTTCTGT 1200
 QY 1358 GGAATGGGCTTACGAGCTGCAACCCCGCAAGTGAACCGTGAAGCCCATCGAGCTTGGCGAG 1417
 Db 1201 GGAATGGGCTTACGAGCTGCAACCCCGCAAGTGAACCGTGAAGCCCATCGAGCTTGGCGAG 1260
 QY 1418 AGGAGAGCTGAGCCGTTGAAAGATCAAGATCAAGAGCTGTGGCAAGTGAATGGGCGAGCC 1477
 Db 1261 AGGAGAGCTGAGCCGTTGAAAGATCAAGATCAAGAGCTGTGGCAAGTGAATGGGCGAGCC 1320
 QY 1478 AGATCTAACCCCGGCTCAAGAGTGGCGCAAGTGTGCAAGCTGTGGCGGCGCGCAAGGCC 1537
 Db 1321 AGATCTAACCCCGGCTCAAGAGTGGCGCAAGTGTGCAAGCTGTGGCGGCGCGCAAGGCC 1380
 QY 1538 TGAACGCAATCGTGGCCCTTGAACGAGAGGCGCGAGCTGAGCTGGCGCGAGAACCGCGAG 1597
 Db 1381 TGAACGCAATCGTGGCCCTTGAACGAGAGGCGCGAGCTGAGCTGGCGCGAGAACCGCGAG 1440
 QY 1598 TCTGTGCGAGGCCGCTGCAAGCGGCTGTACTACAGCCCGCAAGAGACTGTGGCGCGAG 1657
 Db 1441 TCTGTGAGAGGCGCCGCGCAAGAGTGTACTACAGCCCGCAAGAGACTGTGGCGCGAG 1500
 QY 1658 TCCAGAGAGCGGCGCGCAAGAGTGTACTACAGTCTTACCGAGAGCCCTTCAAGAAC 1717
 Db 1501 TCCAGAGAGCGGCGCGCAAGAGTGTACTACAGTCTTACCGAGAGCCCTTCAAGAAC 1560
 QY 1718 TGAAGACCGGCAAGTACGCAAGATGCGCAACCGCGCAACCAAGAGCTGGAAGCAGCTGA 1777
 Db 1561 TGAAGACCGGCAAGTACG 1620
 QY 1778 CCGAGGCGCGTGAAGATCGCAGTGAAGAGCACTGAGTCTGGGCGAAGACCCCAAGT 1837
 Db 1621 CCGAGGCGCGTGAAGATCGCAGTGAAGAGCACTGAGTCTGGGCGAAGATCCCAAGT 1680
 QY 1838 TCGGCTGCGCATTCAGAGAGAGACTGTGGAGACTGTGTGACCACTATGCGAGCGCA 1897
 Db 1681 TCAAGCTGCGCATTCAGAGAGAGACTGTGGAGACTGTGTGAGTATGTGCGAGCGCA 1740
 QY 1898 CTTGATATCCCGAGTGGAGATTGTGTAAACCCCCCTGTGTGAAGCTGTGTGTGTGTGTGT 1957
 Db 1741 CTTGATATCCCGAGTGGAGATTGTGTAAACCCCCCTGTGTGAAGCTGTGTGTGTGTGTGT 1800
 QY 1958 TGGAGAGAGAGCCATTCATCGGCGCGAGACTTTCTTACGTGAGCGGCGCGCGCAACCGCG 2017
 Db 1801 TGGAGAGAGAGCCATTCATCGTGGCGCGAGACTTTCTTACGTGAGCGGCGCGCGCAACCGCG 1860
 QY 2018 AGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGCGCAAGATGTGTAGCC 2077
 Db 1861 AGACCAAGCTGGGCGAGAGCGGCTTACGTGACCGACCGGCGCGCGCGCAAGATGTGTAGCA 1920
 QY 2078 TGAACGAGAGCAACCAAGAGAGCGAGCTGAGAGCACTCAGCTGGGCGCTGAGAGCA 2137
 Db 1921 TGGCGAGAGCAACCAAGAGAGCGAGCTGAGAGCACTCAGCTGGGCGCTGAGAGCA 1980

QY 2138 GCGGAGCGAGGTGAACATCTGTGACCGAGACGCGCAAGTACCGCTTGGGCAATCTCAGGCCC 2197
 Db 1981 GCGGCGTGAAGTGAACATCTGTGACCGAGACGCGCAAGTACCGCTTGGGCAATCTCAGGCCC 2040
 QY 2198 AGCCCGCAAGAGCGGAGCGAGCTGTGAACCAAGTATCTGAGCAGTGTATCAAGAGG 2257
 Db 2041 AGCCCGCAAGAGCGGAGCGAGCTGTGAACCAAGTATCTGAGCAGTGTATCAAGAGG 2100
 QY 2258 AGAAGGTATCTGAGCTGGGTGCGCGCCAGAGGCGCATCGCGCGCAACAGACAGTGC 2317
 Db 2101 AGAAGGTATCTGAGCTGGGTGCGCGCCAGAGGCGCATCGCGCGCAACAGACAGTGC 2160
 QY 2318 ACAAGCTGTGAGCAAGAGGCTATCCGCAAGTGTCTTCTTGAAGCGCATCTGAGCGGCA 2377
 Db 2161 ACAAGCTGTGAGCGCGCGCATCCGCAAGTGTCTTCTTGAAGCGCATCTGAGCGGCA 2220
 QY 2378 TGTGATCTACAGTATCATGAGAGACCTGTAGTGGGCGAGCGGCGCGCTTGAAGTCAATT 2437
 Db 2221 TGTGATCTACAGTATCATGAGAGACCTGTAGTGGGCGAGCGGCGCGCTTGAAGTCAATT 2280
 QY 2438 AAAAGCTTCCGCGGCTAGCACCGGTGAATTC 2469
 Db 2281 AAAAGCTTCCGCGGCTAGCACCGGTGAATTC 2312

RESULT 7
 AX427930 9166 bp DNA 1linear PAT 20-JUN-2002
 LOCUS AX427930
 DEFINITION Sequence 168 from Patent WO0232943.
 ACCESSION AX427930
 VERSION AX427930.1 GI:21538017
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct
 other sequences; artificial sequences.

REFERENCE
 1 Huang, Y. and Nabel, G. J.
 Modifications of hiv env, gag, and pol enhance immunogenicity for
 genetic immunization
 Patent: WO 0232943-A 168 25-APR-2002;
 GOVERNMENT OF THE UNITED STATES (US)

JOURNAL
 Location/Qualifiers

FEATURES
 source
 1..9166
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="plasmid pVR1012x/s containing HIV genes"

ORIGIN
 Query Match 83.1%; Score 2051.2; DB 6; Length 9166;
 Best Local Similarity 92.6%; Pred. No. 3; 6e-168;
 Matches 2192; Conservative 0; Mismatches 163; Indels 13; Gaps 3;

QY 14 TGGCCGAGGCGCATGAGCGCGCAACGCGCAACATCTGTATGCGCGCGCAACTTCA 73
 Db 2961 TGGCCGAGGCGCATGAGCGCGCAACGCGCAACATCTGTATGCGCGCGCAACTTCA 3020
 QY 74 AGGCGCGCGCGCATCATCAAGTGTCTTGAACCTGTGGCGCAAGAGGCGCGCATGCGCGCA 133
 Db 3021 AGGCGCGCGCGCATCATCAAGTGTCTTGAACCTGTGGCGCAAGAGGCGCGCATGCGCGCA 3080
 QY 134 ACTGCGCGCGCGCGCGCAAGAGGCGTGTGAGAGTGTGGCGCAAGAGGCGCGCAACAGATGA 193
 Db 3081 ACTGCGCGCGCGCGCGCAAGAGGCGTGTGAGAGTGTGGCGCAAGAGGCGCGCAACAGATGA 3140
 QY 194 AGGACTGACCGGAGCGCGCGCAAGTCTTCTTCCGCGAGAGCTGTGCTTCTTCCAGGCGCA 253
 Db 3141 AGGACTGACCGGAGCGCGCGCAAGTCTTCTTCCGCGAGAGCTGTGCTTCTTCCAGGCGCA 3199
 QY 254 AGGCGCGCGAGTCTTCCGCGAGAGAGAGACCGGCGCGCAAGCGCGCGCAAGCGCGCGAGCTGC 313
 Db 3200 AGGCGCGAGATTTTCTTCAAGAGAGAGAGAGCGCGCGCAAGCGCGCGCGCGCGCGCGCGCG 3259

QY 314 AGGTGCGCG-----CGAACCCCGCGAGCGCGCGCGAGCGCGCA--- 364
 DB 3260 AGGTTGGGGAAGAGCAACATCTCCTCAAGAGCAGAGCCGATTAACAAGGAACG 3319
 QY 365 ---CCCTGAATCTCCCGCATCACTCTGTGGAGCGCCCTGTGTAGCATCAAGTGG 421
 DB 3320 TATCTTTAGCTTCCCTCAGATCACTTTGGCAGAGACCCCTCTGTCAAAATAAGATGG 3379
 QY 422 GCGGCAGATCAAGAGGCGCTGTGGACAACGCGCGCGAGCAACCTGTGGAGGA 481
 DB 3380 GGGGCGAGCTTGAAGAGGCGCTTCTAGACAACGCGCGCGAGCAACCTGTGGAGGA 3439
 QY 482 TGAAGCTGCGCGCAGAGTGAAGCCCAAGATGATCGCGCGCATCGCGCTTCAACAAG 541
 DB 3440 TGAACCTGCGCGCGCGCTGGAAGCCCAAGATGATCGCGCGCATCGCGCTTCAACAAG 3499
 QY 542 TGGCGCAGTACGACGAGATCTGTATGAGATCTGCGCGCAAGAGCCATCGGACCGTGC 601
 DB 3500 TGGGCGAGTACGACGAGATCTGTATGAGATCTGCGCGCAAGAGCCATCGGACCGTGC 3559
 QY 602 TGAATGCGCGCGCGCGCGCGCGCATGATGCGCGCGCAACCTGTGACCCGAGCTGGGCTGCA 661
 DB 3560 TGGTGGCGCGCGCGCGCGCGCGCATGATGCGCGCGCAACCTGTGACCCGAGATCGCTGCA 3619
 QY 662 CCTGGAATCTCCCGCATCAAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCGCGCATGG 721
 DB 3620 CCTGGAATCTCCCGCATCAAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCGCGCATGG 3679
 QY 722 AGCGGCGCGCAGGTGAAGCGTGGCGCTCTGACGAGAGGAAGATCAAGGCGCTGTACCGCA 781
 DB 3680 AGCGGCGCGCAGGTGAAGCGTGGCGCTCTGACGAGAGGAAGATCAAGGCGCTGTGTGANA 3739
 QY 782 TCTGCGAGGAGTGAAGAGGAGGCGAGATGACCAAGATGCGGCGCGCGAGAACCCCTTACA 841
 DB 3740 TCTGCGAGGAGTGAAGAGGAGGCGAGATGACCAAGATGCGGCGCGCGAGAACCCCTTACA 3799
 QY 842 ACAACCCCGCTGTTCGCGCATCAAGAGAGAGAGCAGCAGCAAGTGGCGCAAGCTGTGTGACT 901
 DB 3800 ACAACCCCGCTGTTCGCGCATCAAGAGAGAGAGCAGCAGCAAGTGGCGCAAGCTGTGTGACT 3859
 QY 902 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTTGAGCTGTGGCATTCGCCACC 961
 DB 3860 TCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGGTTGAGCTGTGGCATTCGCCACC 3919
 QY 962 CGCGCGCGCTGAAGAGAGAGAGGCTGACCGTGTGAGCGTGGCGGAGCGCTACTTCA 1021
 DB 3920 CGCGCGCGCTGAAGAGAGAGAGGCTGACCGTGTGAGCGTGGCGGAGCGCTACTTCA 3979
 QY 1022 GCGTGCCTCTGAGCAGAGACTTCCGCAAGTACACGCGCTTCAACATCCCGCAGATCAACA 1081
 DB 3980 GCGTGCCTCTGAGCAGAGACTTCCGCAAGTACACGCGCTTCAACATCCCGCAGATCAACA 4039
 QY 1082 ACGAGACCCCGCGCATCGCTACAGTACAGTGTGTGCGCCCAAGGCTGTGAAGGCGAGCC 1141
 DB 4040 ACGAGACCCCGCGCATCGCTACAGTACAGTGTGTGCGCCCAAGGCTGTGAAGGCGAGCC 4099
 QY 1142 CGAGCATCTTCAAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCGCAACCCCG 1201
 DB 4100 CGAGCATCTTCAAGAGCAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCGCAACCCCG 4159
 QY 1202 AGATCTGTATCTTACAGTACATGAGCAGCTGTACAGTGTGAGCGAGCGAGCTGTAGATGAGCC 1261
 DB 4160 AGATCTGTATCTTACAGTACATGAGCAGCTGTACAGTGTGAGCGAGCGAGCTGTAGATGAGCC 4219
 QY 1262 AGCAGCTGCGCGCAAGATGAGAGGCTGCGCAGAGCAGCTGTGCGTGGGCTTCAACACCC 1321
 DB 4220 AGCAGCTGCGCGCAAGATGAGAGGCTGCGCAGAGCAGCTGTGCGTGGGCTTCAACACCC 4279
 QY 1322 CGGACAAAGAGCAGAGAGGCGCGCGCTTCTGTGTGAGTGGGCTACAGAGCTGCAACCCCG 1381
 DB 4280 CGGACAAAGAGCAGAGAGGCGCGCGCGCTTCTGTGTGAGTGGGCTACAGAGCTGCAACCCCG 4339

QY 1382 ACAAGTGAACCGTGAAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCGTGAACGACA 1441
 DB 4340 ACAAGTGAACCGTGAAGCCCATCGAGCTGCCGAGAGAGAGAGCTGGAACCGTGAACGACA 4399
 QY 1442 TCCAGAACCTGTGGGCAAGCTGAATCTGGGCGCAGCAAGTCTTACCCCGCATCAAGGTGC 1501
 DB 4400 TCCAGAACCTGTGGGCAAGCTGAATCTGGGCGCAGCAAGTCTTACCCCGCATCAAGGTGC 4459
 QY 1502 GCCAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCATGATCGGACATCGGCGCGTGAACCG 1561
 DB 4460 GCCAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCATGATCGGACATCGGCGCGTGAACCG 4519
 QY 1562 AGAGGCGGAGCTGAGGCTGCGCGCGCGCGCGCGCGCATGATCGGCGCGCGCGCGCGCGCG 1621
 DB 4520 AGAGGCGGAGCTGAGGCTGCGCGCGCGCGCGCGCGCGCATGATCGGCGCGCGCGCGCGCGCG 4579
 QY 1622 TGTACTTACGACCCCGCAGAGAGCTGTATCGCGAGATCGAGAGAGGCGCGCGCGCGCG 1681
 DB 4580 TGTACTTACGACCCCGCAGAGAGCTGTATCGCGAGATCGAGAGAGGCGCGCGCGCGCGCG 4639
 QY 1682 GGAACCTACCAAGTCTTACCAAGAGGCGCTTCAAGAGCTGTGAAGCGCGCAAGTACGCGCAAGA 1741
 DB 4640 GGAACCTACCAAGTCTTACCAAGAGGCGCTTCAAGAGCTGTGAAGCGCGCAAGTACGCGCGCA 4699
 QY 1742 TGGCGACCGCGCGCAGCAGACGATGAGAGCAGCTGACCGAGGCGGTGCGAGAGATCGCGCA 1801
 DB 4700 TGAAGGCGCGCGCAGCAGACGATGAGAGCAGCTGACCGAGGCGGTGCGAGAGATCGCGCA 4759
 QY 1802 TGGAGAGCATGTATCTTGGGCGAGAGCCCGCAAGTTCGCGCTGCGCATTCAGAGAGAGAGA 1861
 DB 4760 CGGAGAGCATGTATCTTGGGCGAGAGCCCGCAAGTTCGCGCTGCGCATTCAGAGAGAGAGA 4819
 QY 1862 CTTGGAGAGCTGTGTGAGCCGACTGAGCAGGCGCGCGCGCGCGCATGATTCGCGAGTGGAGTTGG 1921
 DB 4820 CTTGGAGAGCTGTGTGAGCCGACTGAGCAGGCGCGCGCGCGCGCATGATTCGCGAGTGGAGTTGG 4879
 QY 1922 TGAACACCCCGCGCTGTGTGAAGCTGTGTGATCAAGCTGTGAAGAGAGCCCATCATCGGCG 1981
 DB 4880 TGAACACCCCGCGCTGTGTGAAGCTGTGTGATCAAGCTGTGAAGAGAGCCCATCATCGGCG 4939
 QY 1982 CCGAGACCTTCTAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCATGATTCGCGAGGCGCGCT 2041
 DB 4940 CCGAGACCTTCTAGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCATGATTCGCGAGGCGCGCT 4999
 QY 2042 ACGTGAACGAGCCGCGCGCGCGCGCGAGAGATCGTGAAGCTGTGACCGAGCAGCAGAGAGAGA 2101
 DB 5000 ACGTGAACGAGCCGCGCGCGCGCGCGAGAGATCGTGAAGCTGTGACCGAGCAGCAGAGAGAGA 5059
 QY 2102 CCGAGCTGAGGCGCATTCAGCTGTGCGCTGTGAGAGCAGCGCGCGAGAGGTTGAACATCGTGA 2161
 DB 5060 CCGAGCTGAGGCGCATTCAGCTGTGCGCTGTGAGAGCAGCGCGCGAGAGGTTGAACATCGTGA 5119
 QY 2162 CCGAGACCGAGTACGCGCTGTGGCATATTCAGGCGCGAGCGCGCGCGCGCGAGCGAGCGAGC 2221
 DB 5120 CCGAGACCGAGTACGCGCTGTGGCATATTCAGGCGCGAGCGCGCGCGCGCGCGAGCGAGCGAGC 5179
 QY 2222 TGTGTAAACGAGTATCGAGAGCTGATCAAGAGAGAGAGGTTGAACCTGAAGCTGGGTGC 2281
 DB 5180 TGTGTAAACGAGTATCGAGAGCTGATCAAGAGAGAGAGGTTGAACCTGAAGCTGGGTGC 5239
 QY 2282 CGCGCCAGAGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCATGATCGAGCTGTGTGAGAGAGCATTC 2341
 DB 5240 CGCGCCAGAGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCATGATCGAGCTGTGTGAGAGAGCATTC 5299
 QY 2342 GCAAGGTGTGTTCTGTGACGGCATCGA 2369
 DB 5300 GCAAGGTGTGTTCTGTGACGGCATCGA 5327

RESULT 8
 AX427936
 LOCUS AX427936 9788 bp DNA linear PAT 20-JUN-2002
 DEFINITION Sequence 174 from Patent WO0232943.

ACCESSION AX427936
VERSION AX427936.1 GI:21538023
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Huang, Y. and Nabel, G. J.
TITLE Modifications of hiv env, gag, and pol enhance immunogenicity for
JOURNAL genetic immunization
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
1..9788
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
Query Match 82.7%; Score 2042.8; DB 6; Length 9788;
Best Local Similarity 91.8%; Pred. No. 1.9e-167;
Matches 2197; Conservative 0; Mismatches 177; Indels 18; Gaps 3;
14 TGGCCGAGGCGCATGAGCGGCGCACCGAGCGCAATCTGTATGACGCGGCGAGCACTTGA 73
2961 TGGCCGAGGCGCATGAGCGGCGCACCGAGCGCAATCTGTATGACGCGGCGAGCACTTGA 3020
74 AGGGCCCCAAGGCGCATCATCAATGCTTCACTGCGGCGAAGAGGCGCATGCGCCGCA 133
3021 AGGGCCCCAAGGCGCATCATCAATGCTTCACTGCGGCGAAGAGGCGCATGCGCCGCA 3080
134 ACTGCGCGCGCCCCCGCGAAGAAAGGCTGTGAAAGTGCAGAGGCGCAACGAGATGA 193
3081 ACTGCGCGCGCCCCCGCGAAGAAAGGCTGTGAAAGTGCAGAGGCGCAACGAGATGA 3140
194 AGGACTGCAACCGAGCGCGGCGCACTTCTTCCGCGAAGACCTTGGCTTCCCCAGGGA 253
3141 AGGACTGCAACCGAGCGCGGCGCACTTCTTCCGCGAAGACCTTGGCTTCCCCAGGGA 3194
254 AGGCGCGAGGTTCCCGCAGCGAGAACCGGCGCAAGCGCCCAACGCGCGAGCTGC 313
3195 AGGCGCGAGGTTCCCGCAGCGAGAACCGGCGCAAGCGCCCAACGCGCGAGCTGC 3254
314 AGTGCGCGCG-----CGACAAACCCCGCGAGCGGCGCGCGCGCGCGCGCGCA 364
3255 AGTTTGGGGAAGAGCAACAATCTCTCTCAGAGACGAGGCGCGCGCGCGCGCGCGCG 3314
365 ----CCCTGAATTTCCCGCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCG 421
3315 TATCTTTAGCTTCCCTCAGATCACTTTGGCGAGCGACCCCTCGTCAATTAAGATAG 3374
422 GGGGCGAGATCAAGAGAGGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGAG 481
3375 GGGGCGAGATCAAGAGAGGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGAG 3434
482 TGAAGCTGCGCGCGCAAGTGAATCGGCGCGCATCGGCGCTTCAATCAAG 541
3435 TGAAGCTGCG 3494
542 TGGCGCAGTACGACGAGATCTGTATGAGATCTGCGCGAAGAGCGCGCATCGGCGCTGC 601
3495 TGGCGCAGTACGACGAGATCTGTATGAGATCTGCGCGCAAGAGCGCGCATCGGCGCTGC 3554
602 TGAATGGGCG 661
3555 TGAATGGGCG 3614
662 CCCTGAATTTCCCGCATAGCGCGCATGAGACCGTGCCTGTGAAGCTGAAGCCCGGCGATG 721
3615 CCCTGAATTTCCCGCATAGCGCGCATGAGACCGTGCCTGTGAAGCTGAAGCCCGGCGATG 3674
722 ACGGCG 781

DB 3675 ACGGCG 3734
DB 782 TCTGCGAGAGATGAGAGAGAGGCGGACATCAACGATGCGCCCGAGAACCCCTTACA 841
DB 3735 TCTGCGAGAGATGAGAGAGAGGCGGACATCAACGATGCGCCCGAGAACCCCTTACA 3794
DB 842 ACAACCCCGCTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 3795 ACAACCCCGCTGTGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3854
DB 902 TCCGCGAGCTGAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 961
DB 3855 TCCGCGAGCTGAACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3914
DB 962 CCGCGCGCGCTGAAG 1021
DB 3915 CCGCGCGCGCTGAAG 3974
DB 1022 GGTGCG 1081
DB 3975 GGTGCG 4034
DB 1082 ACGAGACCCCGCGCGCATCGCGCTACAGTCAACAGTCTGCGCGCGCGCGCGCGCGCG 1141
DB 4035 ACGAGACCCCGCGCGCATCGCGCTACAGTCAACAGTCTGCGCGCGCGCGCGCGCGCG 4094
DB 1142 CCAAGCATTTCCAG 1201
DB 4095 CCGCGCATTTCCAG 4154
DB 1202 AGATGATGATCAACAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1261
DB 4155 ACATGATGATCAACAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4214
DB 1262 AGCAGCGCGCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1321
DB 4215 AGCAGCGCGCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4274
DB 1322 CCGAG 1381
DB 4275 CCGAG 4334
DB 1382 ACAAGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCG 1441
DB 4335 ACAAGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCGTGAACCG 4394
DB 1442 TCCAG 1501
DB 4395 TCCAG 4454
DB 4455 GCGAGCTGTGAG 4514
DB 1502 GCGAGCTGTGAG 1561
DB 4515 AGAGAGCGAGCTGTGAG 4574
DB 1622 TGTATCAAG 1681
DB 4575 TGTATCAAG 4634
DB 1682 GGAACCTACAGATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1741
DB 4635 GGAACCTACAGATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4694
DB 1742 TGGCGACGCGCGCAACCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
DB 4695 TGAAGGCG 4754
DB 1802 TGAAGGAGATGATGATCTGCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1861

Db 4755 CCGAGAGCATCTGTGATCTGGGGCGAAGACCCCGAAGTTCAAGCTGCCATCCAGAAAGAGA 4814
Qy 1862 CCTGGAGAGCCTGTGGACCGA CTA CTGGCAGGGCCA CTTGGATCCCGAGTGGAGTTGG 1921
Db 4815 CTTGGAGAGCCTGTGGAGCCGAGTACTGGACAGGCCACTGGATCCCGAGTGGAGTTGG 4874
Qy 1922 TGAACACCCCCCTGTGTGAAGCTGTGTGACCAAGCTGGAAGAAAGGCCATCATCGAGC 1981
Db 4875 TGAACACCCCCCTGTGTGAAGCTGTGTGACCAAGCTGGAAGAAAGGCCATCATCGAGC 4934
Qy 1982 CCGAAGCTTTTCACTGTGAACCGGCGCCGCAACCGGAGACCAAGATCGGCAAGCCGGCT 2041
Db 4935 CCGAAGCTTTTCACTGTGAACCGGCGCCGCAACCGGAGACCAAGCTGTGGCAAGCCGGCT 4994
Qy 2042 AGGTGACCGAACCGGGGCGGCGAAGAAATCGTAGCTGTGACCGAAGACCAACCAAGAGA 2101
Db 4995 AGGTGACCGAACCGGGGCGGCGAAGAAATCGTAGCTGTGACCGAAGACCAACCAAGAGA 5054
Qy 2102 CCGAGCTGAGGCGCATTCAGCTGGCCCTTGACAGACAGCGGAGCGAGGTGAACATCTGTA 2161
Db 5055 CCGAGCTGAGGCGCATTCAGCTGGCCCTTGACAGACAGCGGCGGAGGTGAACATCTGTA 5114
Qy 2162 CCGAAGCCTGTAACCGCTCTGGGATCATTCAGAGCCCAAGCCCAAGAGCGAGCGAGC 2221
Db 5115 CCGAAGCCTGTAACCGCTCTGGGATCATTCAGAGCCCAAGCCCAAGAGCGAGCGAGC 5174
Qy 2222 TGGTGAACGAGATCATTCAGAGCTGTGATCAAGAAAGGATGTACTGTGACTGGTGC 2281
Db 5175 TGGTGAACGAGATCATTCAGAGCTGTGATCAAGAAAGGATGTACTGTGACTGGTGC 5234
Qy 2282 CCGCCCAAGAGGAGCATCGCGGCAACGAGACATCGACCAAGCTGTGACCAAGGCGATTC 2341
Db 5235 CCGCCCAAGAGGAGCATCGCGGCAACGAGACATCGACCAAGCTGTGACCGGCGATTC 5294
Qy 2342 GCAAGGTGCTGTTCTCTGACCGGATCGATGGGCGCATCTGTGATCTAACAGTA 2393
Db 5295 GCAAGGTGCTGTTCTCTGACCGGATCGATGGGCGCATCTGTGATCTAACAGTA 5346

RESULT 9
AX427931 9169 bp DNA linear PAT 20-JUN-2002
LOCUS AX427931
DEFINITION Sequence 169 from Patent WO232943.
ACCESSION AX427931
VERSION AX427931.1 GI:21538018
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS
TITLE
1 Huang Y. and Nabel G.J.
JOURNAL
genetic immunization of hiv env, gag, and pol enhance immunogenicity for
Patent: WO 0232943-A 169 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
1..9169
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"
ORIGIN
Query Match 82.2%; Score 2028.8; DB 6; Length 9169;
Best Local Similarity 91.7%; Pred. No. 3.1e-166;
Matches 2194; Conservative 0; Mismatches 182; Indels 16; Gaps 4;
Qy 14 TGGCGAGGCGCATAGCGCAGCGACCGAGCGCAACATCTGTGATGAGCGCAGCAACTTCA 73
Db 2967 TGGCGAGGCGCATAGCGCAGCGCAGCGACCGAGCGCAACATCTGTGATGAGCGCAGCAACTTCA 3026
Qy 74 AGGGCCCCAAGCGCATCATCAAGTGTCTTCACTGCGGCAAGAGGCGCATCGCCGCA 133
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

Db 3027 AGGGCCAGAAAGCGCAT---CAAGTGTCTTCAACTGCGGCAAGAGAGGCCCACTGGCCGCA 3083
Qy 134 ACTGCGGCGCCCGCCCGCAAGAAAGGCGCTGTGAATTTGGGGCAAGAGAGGCCCAAGAGTA 193
Db 3084 ACTGCGGCGCCCGCCCGCAAGAAAGGCGCTGTGAATTTGGGGCAAGAGAGGCCCAAGAGTA 3143
Qy 194 AGAAGTGAACCGAGGCGCAGGCGCAACTTTTCCGAGAGACTGTGACTTTCCCGAGGCA 253
Db 3144 AGAAGTGAACCGAGGCGCAGGCGCAACTTTTGAAGAGAGATTTGGCTTTCCCAAGAGGA 3202
Qy 254 AGGCGCGAGATTTCACAGCGAGCAACCGGCGCAACAGCCCAACAGCGCGAGCTGC 313
Db 3203 AGGCGCGAGAAATTTCTTCAAGAGCAACAGGCGCAACAGCCCAACAGAAAGAGCTTC 3262
Qy 314 AGGTGCGGCG-----CGAACAACCCCGCAAGGAGCGCGGCGCGAGCGCGCA--- 364
Db 3263 AGGTGCGGAGAAAGAGACAAACATCTCCCTCAAGAGCAGAGGCCAATAGAACAGAACTG 3322
Qy 365 ---CCCTGAACCTTCCCGAGATCACCTGTGGCAGCGCCCGCTGTGAGCATCAAGTGG 421
Db 3323 TATCTTTAGCTTCCCTGAGATCACTTTGGCAGCGACCTCTGTCAATTAAGATAG 3382
Qy 422 GCGGCAGATCAAGAGGCGCTGTGACACCGGCGCGAATCGACACCTGTGAGAGGA 481
Db 3383 GCGGCAGATCAAGAGGCGCTTCTAGACACCGGCGCGAATCGACACCTGTGAGAGGA 3442
Qy 482 TGAAGCTTCCCGGCAAGTGGAAAGCCCAAGATGAGGGGATCGGGCGCTTCAATCAAG 541
Db 3443 TGAAGCTTCCCGGCGCTGGAAGCCCAAGATGAGGGGATCGGGCGCTTCAATCAAG 3502
Qy 542 TGGCGCAATGACACAGATCTGTGATGAGATCTGGGCAAGAAAGGCCATCGGCACTGTC 601
Db 3503 TGGCGCAATGACACAGATCTGTGATGAGATCTGGGCGCAAGAAAGGCCATCGGCACTGTC 3562
Qy 602 TGATGCGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGCTGCA 661
Db 3563 TGATGCGCCCAACCCCGTGAACATCATCGGCGCAACATGCTGACCCAGCTGGCTGCA 3622
Qy 662 CCTGTAACCTTCCCGATCAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 721
Db 3623 CCTGTAACCTTCCCGATCAGCCCGCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 3682
Qy 722 AGGCGCCCAAGGTGAAGAGTGGCCCTGACCCGAGAGAAAGATCAAGGCGCTGACCGCA 781
Db 3683 AGGCGCCCAAGGTGAAGAGTGGCCCTGACCCGAGAGAAAGATCAAGGCGCTGTGAGGA 3742
Qy 782 TCTGCGAGAGATGAGAGAAAGGCGCAAGATCAACAGATCGGCCCGAGAACCTCTTACA 841
Db 3743 TCTGCGAGAGATGAGAGAAAGGCGCAAGATCAACAGATCGGCCCGAGAACCTCTTACA 3802
Qy 842 ACAACCCCGTGTTCGCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGTGAGACT 901
Db 3803 ACAACCCCGTGTTCGCCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGTGAGACT 3862
Qy 902 TCCGAGACTGAAACAAGGCGCACCGAGACTTTGTGGAGGTGACGTGGCATCTCCCGACC 961
Db 3863 TCCGAGACTGAAACAAGGCGCACCGAGACTTTGTGGAGGTGACGTGGCATCTCCCGACC 3922
Qy 962 CCGCGGCGCTGAAGAAAGAAAGAGCGTGAACGCTGTGACGTGGCGCAAGCTTACTTCA 1021
Db 3923 CCGCGGCGCTGAAGAAAGAAAGAGCGTGAACGCTGTGACGTGGCGCAAGCTTACTTCA 3982
Qy 1022 GGTGCGCTTGAACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGACATCAACA 1081
Db 3983 GGTGCGCTTGAACAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGACATCAACA 4042
Qy 1082 ACAGAACCCCGGCGCATCGCTTACAGTAAACAGTCTGCGCCCAAGGAGTGAAGGCGAGCC 1141
Db 4043 ACAGAACCCCGGCGCATCGCTTACAGTAAACAGTCTGCGCCCAAGGAGTGAAGGCGAGCC 4102
Qy 1142 CCAAGCATTTTCAGAGCAGCATGACCAAGATCTTGAAGCCCTTCCGCGCCGCAACCCCG 1201
Db 4103 CCGCATCTTTCAGAGCAGCATGACCAAGATCTTGAAGCCCTTCCGCGCAAGAACCCCG 4162
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

OY	1202	GGATCGGATCTTACAGTAACAATGGACGACCTGTACCTGGGACGACACTGGAGATGGACC	1261
Db	4163	ACATCGATCTTACAGTAACAATGGACGACCTGTACCTGGGACGACACTGGAGATGGACC	4222
OY	1262	AGCAACCGCGCAACATGAGAGAGCTGGCAAGCACTGTGCGTGGGCTTCAACACC	1321
Db	4223	AGCAACCGCGCAACATGAGAGAGCTGGCCAGCACTGTGCGTGGGCTTCAACACC	4282
OY	1322	CCGACAAAGAACACAGAGAGGCCCTCTTCTGTGGATGGGCTACGAGCTGACCCCG	1381
Db	4283	CCGACAAAGAACACAGAGAGGCCCTCTTCTGTGGATGGGCTACGAGCTGACCCCG	4342
OY	1382	ACAAAGTGAACCTGTGCAAGCTCCATCGAGCTGCCGGAAGAGAGCTGGAACCTGTGAACCA	1441
Db	4343	ACAAAGTGAACCTGTGCAAGCTCCATCGTGTCCGGAAGAGACAGTGTGAACCTGTGAACCA	4402
OY	1442	TCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACCCAGATCTTACCCGGGACTCAAGGTGC	1501
Db	4403	TCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACCCAGATCTTACCCGGGACTCAAGGTGC	4462
OY	1502	GCCAGCTGTGCAGACTGTGCGCGCGCCAGAGCCCTGACCGACATGTGCCCCCTGACCG	1561
Db	4463	GCCAGCTGTGCAGACTGTGCGCGCGCCACAGAGCCCTGACCGAGATGTGTCCCCCTGACCG	4522
OY	1562	AGAGAGCCGAGCTGAGAGCTGGCCGAGAACCGCGAGATCTTGCGGAGCCCGTGCACGGC	1621
Db	4523	AGAGAGCCGAGACTGGAAGCTGGCCGAGAACCCGAGATCTTGAGAGAACCCCGTGCACGGC	4582
OY	1622	TGTACTATGACACCCAGCAAGGACTGTGTGGCCGAGATCTCAAGAACAGGAGCCACAGT	1681
Db	4583	TGTACTATGACACCCAGCAAGGACTGTATCGCCGAGATCTCAAGAACAGGAGCCACAGT	4642
OY	1682	GGACCTTACCAATCTTACCAAGAGCCCTTCAAGAACTGAAAGCCGGAAGTACCCCAAGA	1741
Db	4643	GGACCTTACCAATCTTACCAAGAGCCCTTCAAGAACTGAAAGCCGGAAGTACCCCGCA	4702
OY	1742	TGCGCACCGGCCACACCAACGATGGAAGACTGACCGAGGCCGTGCAGAAATGCGCA	1801
Db	4703	TGAAGGGCGGCCACACCAACGATGGAAGACTGACCGAGGCCGTGCAGAAATGCGCA	4762
OY	1802	TGGAGAGCATGTGTGATCTGGGGCAAGAACCCCAAGTTCCGCGCTGCCATCCAGAAAGAGA	1861
Db	4763	CCGAGAGCATGTGTGATCTGGGGCAAGAACCCCAAGTTCCAGCTGCCATCCAGAAAGAGA	4822
OY	1922	TGAACACCCCCCTCTGTGTGAAGCTGTGTGTACCAAGCTGAGAGAGAGCCCATCATCGGCG	1981
Db	4883	TGAACACCCCCCTCTGTGTGAAGCTGTGTGTACCAAGCTGAGAGAGAGCCCATCATCGGCG	4942
OY	1982	CCGAGACCTTCTACGTGACGCGCGCCGCAACCCGCGAGACCAAGATCGGCAAGCCCGCT	2041
Db	4943	CCGAGACCTTCTACGTGACGCGCGCCGCAACCCGCGAGACCAAGCTGGGCAAGCCCGCT	5002
OY	2042	ACGTGACCGAGCCGGGGCCGCGAGAAAGATCTGTGAGCTGTGACCGAGACCAACCAAGAA	2101
Db	5003	ACGTGACCGAGCCGGGGCCGCGAGAAAGTGTGTCCCTTACCGACCAACCAACCAAGAA	5062
OY	2102	CCGAGCTGCAAGCCATTCAGAGCTGGCCCTTGCAAGACAGCGGGACGAGGTGAACTGTGGA	2161
Db	5063	CCGAGCTGCAAGCCATTCACCTTGGCCCTTGCAAGACAGCGGGCTTGAAGGTGAACTGTGGA	5122
OY	2162	CCGACAGCCAGTACGCGCCCTTGGGCATCATTCAGGCCCAAGCCCGACCAAGACGAGACGAGC	2221
Db	5123	CCGACAGCCAGTACCGCCCTTGGGCATCATTCAGGCCCAAGCCCGACCAAGACGAGACGAGC	5182
OY	2222	TGGTGAACCAATCATTCGAGCAGGCTGTGTCAAGAAAGAAAGATGTATCTTGAGCTGGGTGC	2281
Db	5183	TGGTGAACCAATCATTCGAGCAGGCTGTGTCAAGAAAGAAAGATGTATCTTGAGCTGGGTGC	5242

QY	2282	CCGCCCAACAAAGGGCATCGCGCGCAACAGACGATTCGACAAAGCTGTGAGCAAGGCGATCC	2341
Db	5243	CCGCCCAACAAAGGCGATCGCGCGCAACAGACGATTCGACAAAGCTGTGAGCAAGGCGATCC	5302
QY	2342	GCAAGGTGCTGTTCTCGACGGCGCATCGATGGCGGCATCGTGAATTCACAGTA	2393
Db	5303	GCAAGGTGCTGTTCTCGACGGCGCATCGACAAAGGCCGAGAGGACGAGNA	5354
RESULT 10			
AX427925		9194 bp	DNA
LOCUS			linear
DEFINITION	Sequence 163 from Patent WO0232943.		PAT 20-JUN-2002
ACCESSION	AX427925		
VERSION	AX427925.1		GI:21538012
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1..9194			
/organism="synthetic construct"			
/mol_type="unassigned DNA"			
/db_xref="taxon:32630"			
/note="Plasmid pVR1012x/s containing HIV genes"			
ORIGIN			
Query Match	82.1%;	Score 2027;	DB 6;
Best Local Similarity	91.6%;	Pred. No. 4,4e-16;	Length 9194;
Matches 2194;	Conservative 0;	Mismatches 185;	Indels 16;
Gaps 4;			
QY	14	TGGCCGAAGGCATGAGCCAGG---CCACACAGCGCAACATCTCTGATGAGCGCACT	70
Db	2967	TGGCCGAAGGCATGAGCGAGGTGACCAACAGCGCACATCATATGACAGCGCGCACT	3026
QY	71	TCAAGGCGCCCAAGGCGCATCATCAAGTGTTCACCTGCGGCAAGAGGGCCACATCGCC	130
Db	3027	TCCGCAACCAAGGCGCAAGTCGTGAAGTCTTCAACTGCGGCAAGAGGGCCACACCGCC	3086
QY	131	GCAATGCGCGCGCCCGCCGCAAGAGGGCTCTGGAATGCGGCGCAAGAGGCCACCA	190
Db	3087	GCAATGCGCGCGCGCCCGCCGCAAGAGGGCTCTGGAATGCGGCGCAAGAGGGCCACCA	3148
QY	191	TGAAGCATGCAACCGAGCGCCAGGCGCACTTTCGCGAGGACCTTGCTTCCCGCAG	250
Db	3147	TGAAGCATGCAACCGAGCGCGACAGGCTTA-TTTTITTAAGGAAGATCTGCTTCCACAG	3205
QY	251	GCAAGCGCCGCGAGTTCCCGACGACGAAACCGCGCCCAACGCCCCACACGCGCGAGC	310
Db	3206	GGAAGCGCCAGGGAATTTTCTTCAAGCGACGACCAAGGCCACAGCCCCACAGAAAGAGC	3265
QY	311	TGCAAGTGTGCGG-----CGACAAACCCCCCGAGAGACCGCGCGCGGCGGCGACGCA	364
Db	3266	TTCAAGTGTGCGGGAAGACAAACAATCTCTCTCAAGAGCAGAGCCGATATGACAAAGAA	3325
QY	365	-----CCCTGAACCTTCCCGCAGATCACCTGTGGACAGCGCCCCCTGTGTAGCATCAAG	418
Db	3326	CTGTATCTTTTAGTCTTCCCTCAAGTCACTCTTTGGCAGCGACCCCTCGTCAATTAAGA	3388
QY	419	TGGCGGCGCAATCAAGAGGCGCTGTGACACCGCGCGCGACGACACCGTGTGAGG	478
Db	3386	TGGGGGCGCAAGTGAAGAGGCGCTTTCAGACACCGCGCGCGACACCGTGTGAGG	3445
QY	479	AGATGAGCTGCGCGCGCAAGTGAAGGCCCAAGATGATGCGGCGGATCGCGCGCTTCA	538
Db	3446	AGATGAGCTGCGCGCGCGCTGAAGAGCCCAAGATATATGCGGCGGATCGCGCGCTTCA	3505

ORIGIN /note="p1aamid pvr1012x/s containing HIV genes"

Query Match 82.1%; Score 2027; DB 6; Length 12411;
Best Local Similarity 91.6%; Pred. No. 4e-166;
Matches 2194; Conservative 0; Mismatches 185; Indels 16; Gaps 4;

QY 14 TGGCGGAGGCGCATGAGCCAGG---CGACCGGCGCCAACTCTGTATGACGGCCAGCAACT 70
DB 2967 TGGCGGAGGCGCATGAGCCAGGCGCCAACTGTATGAGCGCCAACTGTATGAGCGCCAGCAACT 3026
QY 71 TCAAGGGCGCCCAAGGCGCATCATCAAGTGTCTCAACTGCGGCGCAAGAGGGGCGCATGCGGCC 130
DB 3027 TCCGCAACCAAGCGCAAGATGTGATGAGTGTCTCAACTGCGGCGCAAGAGGGGCGCATCGGCC 3086
QY 131 GCNACTGCGCGCGCGCGCGCGCAAGAGGGGCTGTGAAAGTGTGCGGCAAGAGGGCGCAACAGA 190
DB 3087 GCAACTGCGCGCGCGCGCGCGCAAGAGGGGCTGTGAAAGTGTGCGGCAAGAGGGGCGCAACAGA 3146
QY 191 TGAAGACTGTCAACCGAGCGCGCAAGGCGCACTTCTTCCGCGAGGACTGTGCTTCTCCCGAG 250
DB 3147 TGAAGACTGTCAACCGAGCGCGCAAGGCGCACTTCTTCCGCGAGGACTGTGCTTCTCCCGAG 3205
QY 251 GGAAGCGCGCGCGAGTTCGCCAGCGAGCAACCGGCGCAACGCGCCCAACGCGCGAGC 310
DB 3206 GGAAGCGCGCGAGTTCGCCAGCGAGCAACCGGCGCAACGCGCCCAACGCGCGAGC 3265
QY 311 TGCAGGTGCGCGG-----CGACCAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 364
DB 3266 TTCAGGTTTGGGGAGAGAGCAACAACCTCCCTCTCAGAGAACGAGCGCGCGCGCGCGCGCG 3325
QY 365 -----CCTGAACTTCCCCAGATCACTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 418
DB 3326 CTGTATCTTGTAGCTTCCCTCAGATCACTTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 3385
QY 419 TGGGCGCGCGAGATCAAGGAGGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCG 478
DB 3386 TTAGGGGCGCGAGTCAAGAGGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCG 3445
QY 479 AGATGAGCTGTGCG 538
DB 3446 AGATGAGCTGTGCG 3505
QY 539 AGGTGCGCGAGTACGACGAGTCTGTATGAGATCTGTGCGCGAGAGAGCGCGCATGCGAGCG 598
DB 3506 AGGTGCGCGAGTACGACGAGTCTGTATGAGATCTGTGCGCGAGAGAGCGCGCATGCGAGCG 3565
QY 599 TGTGTATGCG 658
DB 3566 TGTGTATGCG 3625
QY 659 GCACTCTGAATCTTCCCATCAGCGCGCATGAGACGCGTGCCTGTGAAGCTGAAGCGCGGCA 718
DB 3626 GCACTCTGAATCTTCCCATCAGCGCGCATGAGACGCGTGCCTGTGAAGCTGAAGCGCGGCA 3685
QY 719 TGGACGCGCGCGCGAGGAGTGAAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
DB 3686 TGGACGCGCGCGCGAGGAGTGAAGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3745
QY 779 CCATCTGCGAGGAGTGAAGAGAGAGAGAGAGAGTCAACAGATCGCGCGCGCGCGCGCGCG 838
DB 3746 AGATCTGACCGAGAGTGAAGAGAGAGAGAGAGTCAACAGATCGCGCGCGCGCGCGCGCG 3805
QY 839 ACAACACCG 898
DB 3806 ACAACACCG 3865
QY 899 ACTTCCGAGGCTGAACAAGCGCAACCGAGACTTCTTGGAGGCTGCACTGCGCGCGCGCG 958
DB 3866 ACTTCCGAGGCTGAACAAGCGCAACCGAGACTTCTTGGAGGCTGCACTGCGCGCGCGCG 3925
QY 959 ACCCGCGCGCGCGCGAGAGAGAGAGAGCGTGAACCGGTGAGCGTGGGCGAGCGCGCTACT 1018

DB 3926 ACCCGCGCGCGCGCGAGAGAGAGAGAGCGTGAACCGGTGAGCGTGGGCGAGCGCTACT 3985
QY 1019 TCAAGGTGCGCGTGTGAACGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGACATCA 1078
DB 3986 TCAAGGTGCGCGTGTGAACGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGACATCA 4045
QY 1079 ACAACGAGACCG 1138
DB 4046 ACAACGAGACCG 4105
QY 1139 GCCCGAGACTTCTCAGAGCAGCATGACCAAGATCTGTAGCGCTTCCGCGCGCGCGCAAC 1198
DB 4106 GCCCGAGACTTCTCAGAGCAGCATGACCAAGATCTGTAGCGCTTCCGCGCGCGCGCAAC 4165
QY 1199 CCGAGATGTGATCTTACAGATCATGACAGACTTGTAGTGGGCGAGCGCGCGCGCGCGCGCG 1258
DB 4166 CCGAGATGTGATCTTACAGATCATGACAGACTTGTAGTGGGCGAGCGCGCGCGCGCGCGCG 4225
QY 1259 GCGACGACCGCGCGCGAGATGAGAGGCGCGCGAGACCTGTGCGCGTGGGCGTCAACA 1318
DB 4226 GCGACGACCGCGCGAGAGTGAAGAGGCGCGCGAGACCTGTGCGCGTGGGCGTCAACA 4285
QY 1319 CCGCGGACAGAGAGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1378
DB 4286 CCGCGGACAGAGAGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4345
QY 1379 CCGACAGTGAACCGTGTGACCGCATGAGCTGCGCGAGAGAGAGCTGACCGTGAACG 1438
DB 4346 CCGACAGTGAACCGTGTGACCGCATGAGCTGCGCGAGAGAGAGCTGACCGTGAACG 4405
QY 1439 AATCAAGAGTGTGAGGCGCGAGCTGAATCTGGGCGAGCGAGTCAACCGCGCATCAAG 1498
DB 4406 AATCAAGAGTGTGAGGCGCGAGCTGAATCTGGGCGAGCGAGTCAACCGCGCATCAAG 4465
QY 1499 TGCAGCAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1558
DB 4466 TGCAGCAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4525
QY 1559 CCGAGAGGCGCGAGCTGTGAGCTGTGCGCGAGAACCGCGAGATCTGTGCGCGAGCGCG 1618
DB 4526 CCGAGAGGCGCGAGCTGTGAGCTGTGCGCGAGAACCGCGAGATCTGTGCGCGAGCGCG 4585
QY 1619 GCGTGTACTACGACCGCGAGAGAGCTGTGTGGCGGAGTCCAGAGAGCGGCGCAAGAC 1678
DB 4586 GCGTGTACTACGACCGCGAGAGAGCTGTGTGGCGGAGTCCAGAGAGCGGCGCAAGAC 4645
QY 1679 AGTGAACCTACAGAGTCTACAGAGCGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCC 4705
DB 4646 AGTGAACCTACAGAGTCTACAGAGCGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCC 4758
QY 1739 AGATGCGACCG 1798
DB 4706 AGATGCGACCG 4765
QY 1799 CCATGAGAGGACTGTGTGAGTGTGGGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1858
DB 4766 CCATGAGAGGACTGTGTGAGTGTGGGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCG 4825
QY 1859 AGACCTGGAGAGCTGTGTGAACGAGCTACTGTGAAGCGCGCGCGCGCGCGCGCGCGCGCG 1918
DB 4826 AGACCTGGAGAGCTGTGTGAACGAGCTACTGTGAAGCGCGCGCGCGCGCGCGCGCGCGCG 4885
QY 1919 TGTGAAACCG 1978
DB 4886 TGTGAAACCG 4945
QY 1979 GCGCGCGAGCTTCTTACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2038
DB 4946 GCGCGCGAGCTTCTTACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5005
QY 2039 GCTACGTGACCGAGCGCGCGCGCGAGAGATCTGTGAGCTGTGACCGAGACCAACCAACAGA 2098
DB 5006 GCTACGTGACCGAGCGCGCGCGCGAGAGATCTGTGAGCTGTGACCGAGACCAACCAACAGA 5065

QY 2099 AGACGAGCTGACGAGCCATTCAGCTGACCTTGACAGGACGAGCCAGGAGTGAACATCG 2158
DB 5066 AGACGAGCTGACGAGCCATTCAGCTGACCTTGACAGGACGAGCCAGGAGTGAACATCG 5125
QY 2159 TGACCGACGACGAGTACGAGCCCTGAGGATCATTCAGGAGCCAGCCGACAGAGAGAGG 2218
DB 5126 TGACCGACGACGAGTACGAGCCCTGAGGATCATTCAGGAGCCAGCCGACAGAGAGG 5185
QY 2219 AGCTGTGATACGAGATCATTCAGGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGG 2278
DB 5186 AGCTGTGATACGAGATCATTCAGGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGG 5245
QY 2279 TGCCGCGCCACAGAGGACATCGAGCGACAGAGAGATCGACAGAGCTGTGAGCAAGAGCA 2338
DB 5246 TGCCGCGCCACAGAGGACATCGAGCGACAGAGAGATCGAGCTGTGAGCAAGAGCA 5305
QY 2339 TCCGCAAGGTGCTGTCTCTGACAGGATCGATGAGCGAGCATCGTGAATCTACAGTA 2393
DB 5306 TCCGCAAGGTGCTGTCTCTGACAGGATCGATGAGCGAGCATCGTGAATCTACAGTA 5360

RESULT 12
AX427926 9194 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 164 from Patent WO0232943.
DEFINITION AX427926
ACCESSION AX427926
VERSION AX427926.1 GI:21538013
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
TITLE
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 164 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
FEATURES
Location/Qualifiers
1..9194
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR102x/s containing HIV genes"

ORIGIN
Query Match 82.0%; Score 2025.4; DB 6; Length 9194;
Best Local Similarity 91.6%; Pred. No. 6e-166;
Matches 2193; Conservative 0; Mismatches 186; Indels 16; Gaps 4;
QY 14 TGGCCGAGGCGCATGAGCCAGG---CCACGAGGCGCAACATCTGATGACGCGCAACAT 70
DB 2967 TGGCCGAGGCGCATGAGCCAGGAGCAACAGCGCCACATCATGAGCGCGCAACT 3026
QY 71 TCAAGGAGCCCAAGGCGCATCATCAAGTCTTCACTGCGGCAAGAGAGGCGCATGCGCC 130
DB 3027 TCCGCAACGAGGCGCAAGATCGTAGAGTCTTCACTGCGGCAAGAGAGGCGCATCGGCC 3086
QY 131 GCAACTGCGCGCGCCCGCGCAAGAGAGGCTGCTGAAGTGCAGAGAGAGGCGCAACAGA 190
DB 3087 GCAACTGCGCGCGCCCGCGCAAGAGAGGCTGCTGAAGTGCAGAGAGAGGCGCAACAGA 3146
QY 191 TGAAGAGCTGACCGAGGCGCGCAACTTCTTCCGCGAGAGACTGTGCTTCCCGCAG 250
DB 3147 TGAAGAGCTGACCGAGGCGCGCAACTTCTTCCGCGAGAGACTGTGCTTCCCGCAG 3205
QY 251 GCAAGGCGCGCGAGTTCCCGAGAGAGCAACGCGCGCAACAGCGCCACCGCGAGC 310
DB 3206 GCAAGGCGCGCGAGTTCTTCCGAGAGAGCAACAGCGCGCGCGCGCGCGCGCGCAAG 3265
QY 311 TGCAGGTGCGCGG-----CGACAACTCCCGAGCGAGCGAGCGCGCGCGCGCGCGCA 364
DB 3266 TGCAGGTGCGCGAGAGAGCAACAACTCCCTCTCAGAGAGAGAGCGAGCGAGTGAAGAG 3325

QY 365 -----CCTGAACCTTCCCCCAGATCACTGTGTGGCAAGGCCCCCTGTGAGATCAAG 418
DB 3326 CTGTATCTCTTATAGCTTCCCTCAGATCACTTGTGGAGAGACCCCTCGTCAATAAAGA 3385
QY 419 TGGGCGCGCAGATCAAGAGAGGCGCTGTGACACCGGCGCGCAACACCGTCTGAGAG 478
DB 3386 TAGGGGCGCAGCTGAAGAGAGGCGCTTCTTATGACACCGGCGCGCAACACCGTCTGAGAG 3445
QY 479 AGATAGCTCTGCGCGCAAGTGAAGGCCAAGATGATGGCGGCGATTCGGCGCTTCACTCA 538
DB 3446 AGATAGCTCTGCGCGCGCGCTGAGAGGCCAAGATATGGCGGCGATTCGGCGCTTCACTCA 3505
QY 539 AGGTGCGCGAGTACAGACAGATCTGTATCGAGATCTGCGGCGCAAGAGGCGATCGGACCG 598
DB 3506 AGGTGCGCGAGTACAGACAGATCTGTATCGAGATCTGCGGCGCAAGAGGCGATTCGGACCG 3565
QY 599 TGTGTATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGAGCT 658
DB 3566 TGTGTATCGGCGCCACCGCGTGAACATCATCGCGCGCAACATGTGACCCAGATCGGCT 3625
QY 659 GCACTCTGAACCTTCCCATGAGCCCATGAGACCTGTGCGCTGAAAGCTGAAGCCCGCA 718
DB 3626 GCACTCTGAACCTTCCCATGAGCCCATGAGACCTGTGCGCTGAAAGCTGAAGCCCGCA 3685
QY 719 TGGAGGCGCGCCAGAGTGAAGCACTGAGCCCTGACCGAGAGAGATCAAGGCGTGAACCG 778
DB 3686 TGGAGGCGCGCCAGAGTGAAGCACTGAGCCCTGACCGAGAGAGATCAAGGCGTGAACCG 3745
QY 779 CCATCTGCGAGAGATGAGAGAGAGAGGCGCAAGATCGGCGCCCGAGAACCTCT 838
DB 3746 AGATCTGACCGAGATGAGAGAGAGAGGCGCAAGTCAAGCAAGTGGGCGCGAGAACCTCT 3805
QY 839 ACAACACCCCGGTGTTCGCGCATCAAGAGAGACAGCAACAGTGGCGCAAGCTGTGG 898
DB 3806 ACAACACCCCGGTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3865
QY 899 ACTTCGAGAGCTGAACAGAGCGCACCGAATCTTCTGAGAGGTGAGAGCTGAGCATCCCC 958
DB 3866 ACTTCGAGAGCTGAACAGAGCGCACCGAATCTTCTGAGAGGTGAGAGCTGAGCATCCCC 3925
QY 959 ACCCGCGCGCTGAG 1018
DB 3926 ACCCGCGCGCTGAG 3985
QY 1019 TCAAGTGCCTCTGAG 1078
DB 3986 TCAAGTGCCTCTGAG 4045
QY 1079 ACAAGAGAGCCCGCGCATCGCTACAGTGAACAGTGTGCGCGAGAGAGAGAGAGAG 1138
DB 4046 ACAAGAGAGCCCGCGCATCGCTACAGTGAACAGTGTGCGCGAGAGAGAGAGAGAGAG 4105
QY 1139 GCCCGAGATCTTCAAG 1198
DB 4106 GCCCGAGATCTTCAAG 4165
QY 1199 CGAGATGTGTATCTTACAGTACATGAGAGACCTGTGAGAGAGAGAGAGAGAGAGAG 1258
DB 4166 CGAGATGTGTATCTTACAGTACATGAGAGACCTGTGAGAGAGAGAGAGAGAGAGAGAG 4225
QY 1259 GCGACACCGCGCGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318
DB 4226 GCGACACCGCGCGAGAGTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4285
QY 1319 CCCCCGAG 1378
DB 4286 CCCCCGAG 4345
QY 1379 CCGACAGTGAACCTGTGAG 1438
DB 4346 CCGACAGTGAACCTGTGAG 4405


```

Db      ||||| 541 CGGTAAAGCTGAAGCCGCGGATGTAACGCGCCCAAGGTCAAGCATGCGCCCTGACCGAGG 600
Qy      ||||| 758 AGAAGATGAAGGCGCTTGACCGGCATCTTGAGAGAGATGGAAGAGAGGCAAGATCAACA 817
Db      ||||| 601 AGAAGATGAAGGCGCTTGAGAGATCTTGACCGAGATGGAAGAGAGGCAAGATCAACA 660
Qy      ||||| 818 AGATCGGCGCGGAGAACCCCTCAACAACACCCCGGTGTTCGCATCAAGAGAGAGACAGA 877
Db      ||||| 661 AGATCGGCGCGGAGAACCCCTCAACAACACCCCGGTGTTCGCATCAAGAGAGAGACAGA 720
Qy      ||||| 878 CCAAGTGGCGCAAGCTGTGAACTTCCGAGCTGAACAAGCGCAACCCAGAACCTTCTGAG 937
Db      ||||| 721 CCAAGTGGCGCAAGCTGTGAACTTCCGAGCTGAACAAGCGCAACCCAGAACCTTCTGAG 780
Qy      ||||| 938 AGGTCAAGCTGGGAGATCCCGCAACCCCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 997
Db      ||||| 781 AGGTCAAGCTGGGAGATCCCGCAACCCCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 840
Qy      ||||| 998 TGAAGTGGGCGAAGCTCTACTTCAAGGTGCGCTGAGACGAGACTTCCGCAAGTACAGG 1057
Db      ||||| 841 TGAAGTGGGCGAAGCTCTACTTCAAGGTGCGCTGAGACGAGACTTCCGCAAGTACAGG 900
Qy      ||||| 1058 CTTTCAACATCCCGACAGATCAACAACGAGACCCCGGCAATCCGCTACCAAGTACAGG 1117
Db      ||||| 901 CTTTCAACATCCCGACAGATCAACAACGAGACCCCGGCAATCCGCTACCAAGTACAGG 960
Qy      ||||| 1118 TGCCTCAGGCGTGGAGAGGCGACCCCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1177
Db      ||||| 961 TGCCTCAGGCGTGGAGAGGCGACCCCGAGATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1020
Qy      ||||| 1178 AGCCCTTCGCGCGCGCAACCCCGAGATCTGATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1237
Db      ||||| 1021 AGCCCTTCGCGCGCGCAACCCCGAGATCTGATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1074
Qy      ||||| 1238 TGGGCGAGCGACTTGAGAGATGCGCCAGACCGCGCGCAAGATGAGAGAGAGAGAGAGAG 1297
Db      ||||| 1075 TGGGCGAGCGACTTGAGAGATGCGCCAGACCGCGCGCAAGATGAGAGAGAGAGAGAG 1134
Qy      ||||| 1298 TGCCTCAGGCGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db      ||||| 1135 TGCCTCAGGCGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1194
Qy      ||||| 1358 GATGCGGCTACAGAGCTGACACCCCGCAAGTGAACCGTGAAGCCATCGAGCTGCGCGAGA 1417
Db      ||||| 1195 GATGCGGCTACAGAGCTGACACCCCGCAAGTGAACCGTGAAGCCATCGAGCTGCGCGAGA 1254
Qy      ||||| 1418 AGAGAGCTGGAACCGTGAACGATTCAGAGAGCTGTGGGCAAGCTGAATCTGGGCGAGCC 1477
Db      ||||| 1255 AGAGAGCTGGAACCGTGAACGATTCAGAGAGCTGTGGGCAAGCTGAATCTGGGCGAGCC 1314
Qy      ||||| 1478 AGATCAACCCCGGCAATCAAGGTGCGCAAGCTGTGAAGAGTGTGCGCGCGCGCGAGAGCC 1537
Db      ||||| 1315 AGATCAACCCCGGCAATCAAGGTGCGCAAGCTGTGAAGAGTGTGCGCGCGCGCGAGAGCC 1374
Qy      ||||| 1538 TGAACCGACATCGTGCCTGCTGACCGAGAGAGCGAGCTGAGAGTGGCGAGAGAGAGAGAG 1597
Db      ||||| 1375 TGAACCGAGTGAATCCCTGCTGACCGAGAGAGCGAGCTGAGAGTGGCGAGAGAGAGAGAG 1434
Qy      ||||| 1598 TCCCTGCGAGAGCCCGCTGCAACGCGGTGTACTAGAGACCCCGAGAGAGAGAGAGAGAGAG 1657
Db      ||||| 1435 TCCCTGAGAGAGCCCGCTGCAACGAGGTGTACTAGAGACCCCGAGAGAGAGAGAGAGAGAG 1494
Qy      ||||| 1658 TCCAGAGAGAGGCGCAACGAGCTGAGACTTACAGATCTTACAGAGAGAGAGAGAGAGAGAG 1717
Db      ||||| 1495 TCCAGAGAGAGGCGCAACGAGCTGAGACTTACAGATCTTACAGAGAGAGAGAGAGAGAGAG 1554
Qy      ||||| 1718 TGAAGACCGGCAAGTACGCAAGATGCGACCGCGCAACGAGAGAGAGAGAGAGAGAGAGAG 1777
Db      ||||| 1555 TGAAGACCGGCAAGTACGCGCGCGCATGCGGCGCGCGCAACGAGAGAGAGAGAGAGAGAGAG 1614
Qy      ||||| 1778 CCGAGGCGGTGAGAGAGATGCGCATGAGAGAGATCTGTGAGGCGAGAGAGAGAGAGAGAG 1837

```

```

Db      ||||| 1615 CCGAGGCGGTGAGAGAGGTGAGACCGAGAGCATGTGATCTGGGCAAGATCCCAAGT 1674
Qy      ||||| 1838 TCCGCTCCCTCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1897
Db      ||||| 1675 TCAAGCTGCCCATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
Qy      ||||| 1898 CTTGATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1957
Db      ||||| 1735 CTTGATCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Qy      ||||| 1958 TGAAGAGAGAGCCCATCATCGGCGCGAGAGACTTCTTACGTGAGACGCGCGCGCAACCGCG 2017
Db      ||||| 1795 TGAAGAGAGAGCCCATCATCGTGGCGCGCGAGAGACTTCTTACGTGAGACGCGCGCGCAACCGCG 1854
Qy      ||||| 2018 AGACCAAGATCGGCAAGGCGCGCTTACGTGAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2077
Db      ||||| 1855 AGACCAAGCTGGGCAAGGCGCGCTTACGTGAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Qy      ||||| 2078 TGAACGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2137
Db      ||||| 1915 TGCAGGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
Qy      ||||| 2138 GCGGCAAGAGGTGAACATCTGTACCGACAGCCAGTACGCGCTGGGCAATCATCAAGGCC 2197
Db      ||||| 1975 GCGGCTGAGAGGTGAACATCTGTACCGACAGCCAGTACGCGCTGGGCAATCATCAAGGCC 2034
Qy      ||||| 2198 AGCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2257
Db      ||||| 2035 AGCCCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2094
Qy      ||||| 2258 AGAAGGTGATCTGAGTGTGAGTCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2317
Db      ||||| 2095 AGAAGGTGATCTGAGTGTGAGTCCCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2154
Qy      ||||| 2318 ACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2377
Db      ||||| 2155 ACAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2214
Qy      ||||| 2378 TCGTATCTTACAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2437
Db      ||||| 2215 TCGTATCTTACAGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2274
Qy      ||||| 2438 AAAAGCTTCCGCGGCTAGACCGGTGAATTC 2469
Db      ||||| 2275 AAAAGCTTCCGCGGCTAGACCGGTGAATTC 2306

```

```

RESULT 14
CO870574      2306 bp      DNA      linear      PAT 13-SEP-2004
LOCUS        CO870574
DEFINITION   Sequence 82 from Patent EP1433851.
ACCESSION    CO870574
VERSION      CO870574.1 GI:5200090
KEYWORDS
SOURCE
  ORGANISM   synthetic construct
  SOURCE     synthetic construct

```

REFERENCE 1 other sequences; artificial sequences.

AUTHORS Barnett,S., Zurnegede,U., Srivastava,I., Llan,Y., Hartog,K., Liu,H., Greer,C., Selby,M. and Walker,C.
 TITLE Improved expression of HIV polypeptides and production of virus-like particles
 JOURNAL Patent: RP 1433851-A 82 30-JUN-2004;
 CHIRON CORPORATION (US)

FEATURES
 source location/Qualifiers
 1..2306

/organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Description of Artificial Sequence:
 FS(-).promod.RTopt.YM"

ORIGIN

Query Match 82.0%; Score 2025.2; DB 6; Length 2306;
Best Local Similarity 93.3%; Pred. No. 1e-165; Indels 18; Gaps 3;
Matches 2156; Conservative 0; Mismatches 138;

170 GCGGCAAGAGGCGCAACAGATGAAAGACTGCAACCGAGCGCAAGCTTCTTCCGCG 229
1 GGGGCGGGAAGGACCAATTAAGATTGCACTGAGAGACAGGTATTCTTCCGCG 60
230 AGGACTGGCTTCTCCCAAGGCGAAGGCCCGGAGTTCCCAAGAGCAAAACCGCGCA 289
61 AGGACTGGCTTCTCTGAGAGGCGAAGGCCCGGAGTTCAAGCGAGAGACCGCGCGCA 120
290 AAGAGCCCAACAGCCGCGAGCTGAGAGTCCGCGCG-----AAGACCCCGCAGAGAG 343
121 AAGAGCCCAACAGCCGCGAGCTGAGAGTGTGGGCGCGAAGAACAAACCTGAGAGAG 180
344 CCGGCGCGAGGCGCAAGGCAACCTG-----AATTCCCGAGATCAACCTGTGACG 397
181 CCGGCGCGAACCAGCGGCAACCTGAGCTTCAACTTCCCGAGATCAACCTGTGACG 240
398 GCGCCCTGTGAGCATCAAGTGGGCGGCAAGTCAAGAGAGCCCTGTGACACCGGCG 457
241 GCGCCCTGTGAGCATCAAGATCGGCGGCAAGCTCAAGAGAGCGCTGACACCGGCG 300
458 CCGAGCAACCTGTGTGAGAGAGATGAGCTTCCCGCGCAAGTGAAGCCCAAGATATCG 517
301 CCGAGCAACCTGTGTGAGAGAGATGAGCTTCCCGCGCAAGTGAAGCCCAAGATATCG 360
518 GCGGCACTGGGCGCTTCAATCAAGGTGCGCAAGTCAAGAGATCTGTATGAGATCTGCG 577
361 GCGGCACTGGGCGCTTCAATCAAGGTGCGGCAAGTCAAGAGATCTCCGTGAGATCTGCG 420
578 GCAAGAGGCGCATCGGCAACCTGTGATCGGCGGCAACCCCGGTGAACATCATCGGCGCA 637
421 GCAAGAGGCGCATCGGCAACCTGTGATCGGCGGCAACCCCGGTGAACATCATCGGCGCA 480
638 ACATGTGAGCCAGCTGGGCTGTGACCTTGAATTTCCCATCAAGCCCATCTGAGACCTGTG 697
481 ACCTGTGAGCCAGATCGGCTGTGACCTTGAATTTCCCATCAAGCCCATCTGAGACCTGTG 540
698 CCGTGAAGCTGAAGCCCGGATGGAAGGCGCCCAAGGTGAAGAGCTGGCCCTGACCGGAG 757
541 CCGTGAAGCTGAAGCCCGGAGATGGAAGGCGCCCAAGGTGAAGAGCTGGCCCTGACCGGAG 600
758 AGAAGATCAAGGCGCTGTACCGCCATCTGCGAGAGATGGAAGAGAGGCGCAAGATCAACA 817
601 AGAAGATCAAGGCGCTGTGTGAGATCTGACCGAGATGGAAGAGAGGCGCAAGATCAACA 660
818 AGATCGGCGCCCGAGAACCCCTTCAACAACCCCGGTGTTCCCATCAAGAGAGAGACACA 877
661 AGATCGGCGCCCGAGAACCCCTTCAACAACCCCGGTGTTCCCATCAAGAGAGAGACACA 720
878 CCAAGTGGGCGCAAGCTGTGTGAGATTTCCGCGAGGTGAACAAGCGCAACCGAGATTTCTGG 937
721 CCAAGTGGGCGCAAGCTGTGTGAGATTTCCGCGAGGTGAACAAGCGCAACCGAGATTTCTGG 780
938 AGGTGAGCTGGGCAATCCCGACCCCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 997
781 AGGTGAGCTGGGCAATCCCGACCCCGCGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 840
998 TGAAGCTGGGCGAGCGCTTCAAGCGTGTCCCTTGAACGAGAGCTTCCGCAAGTACACCG 1057
841 TGAAGCTGGGCGAGCGCTTCAAGCGTGTCCCTTGAACGAGAGCTTCCGCAAGTACACCG 900
1058 CCTTCAACATCCCAAGATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1117
901 CCTTCAACATCCCAAGATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
1118 TGGCCAGAGGCTGAAAGGAGCGCCAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1177
961 TGGCCAGAGGCTGAAAGGAGCGCCAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAG 1020
1178 AGCCCTTCCGCGCGCGCAACCCGAGATGTGATCTAACAAGTCAATGAGAGAGAGAGAG 1237

1021 AGCCCTTCCGAGAGAGAACCCCGACATCTGTATCTTACCA-----GGCCCCCTGTAGC 1074
1238 TGGGAGGCACTTGAAGATCGGCGAGCAACCGCGCAAGATCGAGAGAGCTGCGCAAGACC 1297
1075 TGGGAGGCACTTGAAGATCGGCGAGCAACCGCGCAAGATCGAGAGAGCTGCGCGAGACC 1134
1298 TGTCTGCTGGGCGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
1135 TGTCTGCTGGGCGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1194
1358 GGAATGGGCTACGAGCTGACCCCGGCAAGTGAACGCTGAGAGCCATTCAGAGCTGCCGAGA 1417
1195 GGAATGGGCTACGAGCTGACCCCGGCAAGTGAACGCTGAGAGCCATTCAGAGCTGCCGAGA 1254
1418 AGAAGAGCTGACCGTGAACGACATCTCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCC 1477
1255 AGAAGAGCTGACCGTGAACGACATCTCAAGAGCTGTGGCAAGCTGAACCTGGCGAGCC 1314
1478 AGATCTAACCCCGGCAATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCC 1537
1315 AGATCTAACCCCGGCAATCAAGTGAAGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCC 1374
1538 TGAACGACATCGTGTCCCTTGAACGAGAGAGCGGAGCTGAGAGCTGGCGGAGAAACCGGAGA 1597
1375 TGAACGAGGTATTCCTCTGACCGAGAGAGCGGAGCTGAGAGCTGGCGGAGAAACCGGAGA 1434
1598 TCTCTGCGAGGCGCGTGTGACGCGCTGTACTAAGACCCCGAGCAAGAGCTGTGTGCGGAGA 1657
1435 TCTCTGAGAGAGCGCGTGTGACGAGGTGTACTAAGACCCCGAGCAAGAGCTGTGTGCGGAGA 1494
1658 TCCAGAGAGAGGCGCAACGACGTGAGCTTACAGATCTAACAAGAGGCTTCAAGAGACC 1717
1495 TCCAGAGAGAGGCGCGAGGCGAGTGAACCTTCAAGATCTAACAAGAGGCTTCAAGAGACC 1554
1718 TGAAGAGCGGCGAGTACGCAAGATGGCGCACCGGCCCAACCAACGAGCTGAAGAGAGCTGA 1777
1555 TGAAGAGCGGCGAGTACGCGGCGGAGCGGCGCCCAACCAACGAGCTGAAGAGAGCTGA 1614
1778 CCGAGGCGCTGTGAGAGATCGCCATGTAGAGAGCATGTGATCTTGGGCGAAGACCCCAAGT 1837
1615 CCGAGGCGCTGTGAGAGAGTGAACCGAGAGCATGTGATCTTGGGCGAAGATCCCAAGT 1674
1838 TCCGCTGTCCATCAAGAGAGAGAGCTTGGAGAGCTGTGTGAACCGACTTCTGCGAGCGCA 1897
1675 TCAAGCTGTCCATCAAGAGAGAGAGCTTGGAGAGCTGTGTGAAGTGTGCTGCGAGCGCA 1734
1898 CCTGATCCCGGAGTGGAGGTTCTGTGAACAACCCCGGCTGTGTGAAGCTGTGTGACGAG 1957
1735 CCTGATCCCGGAGTGGAGGTTCTGTGAACAACCCCGGCTGTGTGAAGCTGTGTGACGAG 1794
1958 TGAAGAGAGAGGCGCATCATCGGCGCGAGAGCTTCTAAGTGAACGCGCGCGCAACCGCG 2017
1795 TGAAGAGAGAGGCGCATCATCGGCGCGAGAGCTTCTAAGTGAACGCGCGCGCAACCGCG 1854
2018 AGAACAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGCAAGAGATGTGAGCC 2077
1855 AGAACAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGCAAGAGATGTGAGCA 1914
2078 TGAACGAGACCAACCAAG 2137
1915 TGAACGAGACCAACCAAG 1974
2138 GCGGAGAGAGGCTGAACATCGTGAACGAGCAGCATCGCTTGGGCAATTCAGAGCGCC 2197
1975 GCGGAGAGAGGCTGAACATCGTGAACGAGCAGCATCGCTTGGGCAATTCAGAGCGCC 2034
2198 AGCCCGAACAAG 2257
2035 AGCCCGAACAAG 2094
2258 AGAAGGTATCTGAGCTGGGCGCGCGCAAGGCGCATCGGCGCGCAAGAGAGAGAGAGAG 2317

Db 1555 TGAAGACCGGCAAGTACGCCCCGATCGCGGCGCCCAACCAACGAGTGAAGCACTGA 1614
QY 1778 CCGAGGCCGTGCAAGAGATCGCCATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGT 1837
Db 1615 CCGAGGCCGTGCAAGAGATGAGACCGAGAGCATCTGTATCTGGGGCAAGATCCCAAGT 1674
QY 1838 TCGGCTTGCCCATCCAGAGAGAGACTGGGAGACTGTGTGAACCGACTACTGGCAGGCCA 1897
Db 1675 TCAAGCTGCCATCCAGAGAGAGACTGGGAGGCTGTGTGAATGAGTACTGGCAGGCCA 1734
QY 1898 CCGATCCCGAGTGGAGATTGTGAACACCCCCCGTGTGAAGCTGTGTACGAGC 1957
Db 1735 CTTGATCCCGAGTGGAGATTGTGAACACCCCCCGTGTGAAGCTGTGTACGAGC 1794
QY 1958 TGAAGAGAGGCCCATCATCGGCGCCGAGACCTTCTACGTGACGGCGCCGCCACCGCG 2017
Db 1795 TGAAGAGAGGCCCATCTGTGGCGCCGAGACCTTCTACGTGACGGCGCCGCCACCGCG 1854
QY 2018 AGACCAAGATCGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAAGAGTGTGAGCC 2077
Db 1855 AGACCAAGCTGGGCAAGGCCGGCTACGTGACCGACCGGGGCCGGCAAGAGTGTGAGCA 1914
QY 2078 TGAACCGAGCCCAACCAAGAGAGACCGAGCTGAGGCCATCCAGCTGGCCCTGCAAGACA 2137
Db 1915 TCGCCGACACACCAACCAAGAGACCGAGCTGAGGCCATCCAGCTGGCCCTGCAAGACA 1974
QY 2138 GCGGCAAGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2197
Db 1975 GCGGCTGTGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTGGGCATCATCCAGGCC 2034
QY 2198 AGCCCGACAAAGACGAGAGCTGTGAACCAATCATCGAGCAGCTGATCAAGAGG 2257
Db 2035 AGCCCGACAAAGACGAGAGCTGTGAACCAATCATCGAGCAGCTGATCAAGAGG 2094
QY 2258 AGAAGGTGTACCTGAGCTGGGTGCCGCCCAAGAGGCATCGGGGCAACGAGCAGATCG 2317
Db 2095 AGAAGGTGTACCTGGCTGGGTGCCGCCCAAGAGGCATCGGGGCAACGAGCAGGTGG 2154
QY 2318 ACNAGCTGTGAGCAAGGCGATCCGCAAGGTGCTGTTCTGTGACGGCATCATGAGCGCA 2377
Db 2155 ACNAGCTGTGAGCGCGCGCATCCGCAAGGTGCTGTTCTGTGACGGCATCATGAGCGCA 2214
QY 2378 TCGTGAATCTACAGTACATGAGCAGCTGTACGTGGGCAAGCGGCGCTTAGATCGATT 2437
Db 2215 TCGTGAATCTACAGTACATGAGCAGCTGTACGTGGGCAAGCGGCGCTTAGATCGATT 2274
QY 2438 AAAAGCTTCCCGGGGCTAGACCGGTGAATTTC 2469
Db 2275 AAAAGCTTCCCGGGGCTAGACCGGTGAATTTC 2306

Search completed: December 30, 2005, 19:16:16
Job time : 11995.5 secs

This Page Blank (uspro)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1299.83 Seconds
(without alignment(s))
12659.489 Million cell updates/sec

Title: US-09-610-313B-30

Perfence score: 2469
Sequence: 1 gtcgacgccaccatgagccga.....gggctagcaccggtgaattc 2469

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21:*

- 1: geneseqn1980a:*
- 2: geneseqn1990a:*
- 3: geneseqn2000a:*
- 4: geneseqn2001a:*
- 5: geneseqn2001a:*
- 6: geneseqn2002a:*
- 7: geneseqn2002a:*
- 8: geneseqn2003a:*
- 9: geneseqn2003a:*
- 10: geneseqn2003a:*
- 11: geneseqn2003a:*
- 12: geneseqn2004a:*
- 13: geneseqn2004a:*
- 14: geneseqn2005a:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2469	100.0	2469	6	ABL39959
2	2469	100.0	2469	12	ADM73764
3	2457	99.5	2457	8	ACA03548
4	2457	99.5	2457	10	ADCl3266
5	2442.2	98.9	2453	6	ABL39960
6	2442.2	98.9	2463	12	ADM73765
7	2434.6	98.6	2457	8	ACA03547
8	2434.6	98.6	2457	10	ADCl3265
9	2415.4	97.8	2457	6	ABL39961
10	2415.4	97.8	2457	12	ADM73766
11	2401.8	97.3	2445	8	ACA03546
12	2401.8	97.3	2445	10	ADCl3264
13	2394.8	97.0	3930	10	ADCl3230
14	2393.2	96.9	3930	10	ADCl3231
15	2393.2	96.9	3930	10	ADCl3232
16	2393.2	96.9	5184	8	ACA03591
17	2393.2	96.9	5184	10	ADCl3279
18	2362.8	95.7	3531	10	ADCl3234
19	2361.2	95.6	3537	10	ADCl3236

20	2360.2	95.6	5145	8	ACA03521	ACA03521 Synthetic
21	2360.2	95.6	5145	10	ADCl3233	ADCl3233 DNA of HI
22	2350.2	95.2	3538	10	ADCl3235	ADCl3235 DNA of HI
23	2349.4	95.2	3624	8	ACA03550	ACA03550 Synthetic
24	2349.4	95.2	3624	10	ADCl3268	ADCl3268 DNA of HI
25	2301.8	93.2	3607	8	ACA03551	ACA03551 Synthetic
26	2301.8	93.2	3607	10	ADCl3269	ADCl3269 DNA of HI
27	2283.6	92.5	3597	10	ACA03549	ACA03549 Synthetic
28	2283.6	92.5	3597	10	ADCl3267	ADCl3267 DNA of HI
29	2149.6	87.1	2472	8	ACA03543	ACA03543 Synthetic
30	2149.6	87.1	2472	8	ACC78507	ACC78507 HIV p2Pol
31	2121.2	85.9	2466	8	ACA03542	ACA03542 Synthetic
32	2121.2	85.9	2466	8	ACC78506	ACC78506 HIV p2Pol
33	2094.4	84.8	2460	8	ACA03541	ACA03541 Synthetic
34	2094.4	84.8	2460	8	ACC78505	ACC78505 HIV p2Pol
35	2093.4	84.8	3564	8	ACC78488	ACC78488 HIV GagPo
36	2093.4	84.8	3564	8	ACC78489	ACC78489 HIV GagPo
37	2092.8	84.8	4716	8	ACA03522	ACA03522 Synthetic
38	2092.8	84.8	4716	10	ADCl3238	ADCl3238 DNA of HI
39	2089	84.6	3999	8	ACC78484	ACC78484 HIV GagCo
40	2087.4	84.5	3999	8	ACC78485	ACC78485 HIV GagCo
41	2087.4	84.5	3999	8	ACC78486	ACC78486 HIV GagCo
42	2087.4	84.5	5283	8	ACA03584	ACA03584 Synthetic
43	2087.2	84.5	5283	8	ACC78529	ACC78529 HIV TatRe
44	2087.2	84.5	4713	8	ACA03592	ACA03592 Synthetic
45	2087.2	84.5	4713	10	ADCl3280	ADCl3280 DNA of HI

ALIGNMENTS

RESULT 1	ABL39959	standard; DNA; 2469 BP.
ID	ABL39959	
XX	ABL39959;	
AC		
XX		
DT	15-MAY-2002 (first entry)	
XX		
DE	Synthetic construct PR975(+) SEQ ID NO:30.	
XX		
KW	Human immunodeficiency virus type C; antigenic HIV type C protein;	
KW	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;	
KW	immunostimulant; gene therapy; gene; ds.	
OS	Human immunodeficiency virus; type C.	
XX	Synthetic.	
XX		
FN	WO200204493-A2.	
XX		
PD	17-JAN-2002.	
XX		
PF	05-JUL-2001; 2001WO-US021241.	
XX		
PR	05-JUL-2000; 2000US-00610313.	
XX		
PA	(CHIR) CHIRON CORP.	
XX	(UYST-) UNIV STELLENBOSCH.	
XX		
PI	Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	
XX		
DR	WPI; 2002-154920/20.	
XX		
PT	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	
PT	applications including DNA immunization or generation of packaging cell	
PT	lines, particularly in gene therapy.	
XX		
PS	Claim 1; Fig 8; 233p; English.	
XX		
CC	The present invention describes expression cassettes comprising a	
CC	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	
CC	type C polypeptides. The expression cassettes comprise any of the HIV	
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	

CC (1). (1) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX

SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;

Query Match 100.0%; Score 2469; DB 6; Length 2469;
Best Local Similarity 100.0%; Pred. No. 8,3e-298;

Matches 2469; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGACGACCACTGACGAGGCGCATGAGCCAGGCGCCAACTCGATGACG 60
DB 1 GTGACGACCACTGACGAGGCGCATGAGCCAGGCGCCAACTCGATGACG 60
QY 61 CGCAGCACTTCAAGGGGCGCCAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCG 120
DB 61 CGCAGCACTTCAAGGGGCGCCAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCG 120
QY 121 CACATGCGCGGCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
DB 121 CACATGCGCGGCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
QY 181 GGCACACAGATGAGGAGCTGACACGAGCGCGAGCGCACTTCTCGCGAGGAGCTGCGC 240
DB 181 GGCACACAGATGAGGAGCTGACACGAGCGCGAGCGCACTTCTCGCGAGGAGCTGCGC 240
QY 241 TTCCCGCAAGGAGGAGCGCGAGGTTCCCGAGCGAGCAGACCGCGCCAAAGCGCCACC 300
DB 241 TTCCCGCAAGGAGGAGCGCGAGGTTCCCGAGCGAGCAGACCGCGCCAAAGCGCCACC 300
QY 301 ACCCGGAGCTGACAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
DB 301 ACCCGGAGCTGACAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 361 GGCACCGTGAATCTTCCCGAGATCACTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCG 420
DB 361 GGCACCGTGAATCTTCCCGAGATCACTGTGAGCAGCGCGCGCGCGCGCGCGCGCGCG 420
QY 421 GCGCGGCAAGATCAAGAGGCGCGTGTGAGCACCGCGCGCGAGCAGACCGTGTGAGAG 480
DB 421 GCGCGGCAAGATCAAGAGGCGCGTGTGAGCACCGCGCGCGAGCAGACCGTGTGAGAG 480
QY 481 ATGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
DB 481 ATGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 541 GTGCGGCAATGACGACGATCTCGATGAGATCTGTGCGGAGAGGCGCATTCGGGCAAC 600
DB 541 GTGCGGCAATGACGACGATCTCGATGAGATCTGTGCGGAGAGGCGCATTCGGGCAAC 600
QY 601 CTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
DB 601 CTGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
QY 661 ACCCTGAATCTTCCCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
DB 661 ACCCTGAATCTTCCCGATCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
DB 721 GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
QY 781 ATCTGCGAGAGATGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
DB 781 ATCTGCGAGAGATGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 841 AACACCGCGCGTGTGCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AACACCGCGCGTGTGCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
```

```
QY 901 TTCCCGAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 901 TTCCCGAGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
QY 961 CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
DB 961 CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
QY 1021 AGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
DB 1021 AGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 1081 AACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
DB 1081 AACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1141 CCCAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1141 CCCAGCATCTTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1201 GAGATGTGATTTACAGATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1201 GAGATGTGATTTACAGATGACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1261 GAGACCGCGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
DB 1261 GAGACCGCGCGCGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1321 CCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 CCCGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 GACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 GACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 ATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1441 ATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 GCGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 GCGCAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 GTGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 GTGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TGGACCTTACAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1681 TGGACCTTACAGATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 ATGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800
DB 1741 ATGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1800
QY 1801 ATGAGAGAGATGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 ATGAGAGAGATGTGATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 ACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
DB 1861 ACTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1921 GTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
DB 1921 GTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
```


Db 721 GACGGCCCCAAGGTGAAGCATGTGGCCCTGACCGAGAGAAAGATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGCGAGAGATGGAAGAAAGAGGCGCAAGATCAACAAATGGGCCCGGAGAACCCCTAC 840
Db 781 ATCTGCGAGAGATGGAAGAAAGAGGCGCAAGATCAACAAATGGGCCCGGAGAACCCCTAC 840
Qy 841 AACACCCCGGTGTGCGCATCAAGAAAGAGACAGACCAAGTGGCGCAAGCTGTGTGAC 900
Db 841 AACACCCCGGTGTGCGCATCAAGAAAGAGACAGACCAAGTGGCGCGCAAGCTGTGTGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGTGAGGTGACGTGGGCAATCCCCAC 960
Db 901 TTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGTGAGGTGACGTGGGCAATCCCCAC 960
Qy 961 CCCGCGGCTGAAGAAAGAAAGAGCGTGAACGCTGTGAGAGCTGTGGGCAATCCCTTC 1020
Db 961 CCCGCGGCTGAAGAAAGAAAGAGCGTGAACGCTGTGAGAGCTGTGGGCAATCCCTTC 1020
Qy 1021 AACGTGCCCCGTGACGAGACTTCCGCAAGTACACCGCTTGCACATCCCGACAGATCAAC 1080
Db 1021 AACGTGCCCCGTGACGAGACTTCCGCAAGTACACCGCTTGCACATCCCGACAGATCAAC 1080
Qy 1081 AACGAGACCCCGGCAATCGGCTACCAAGTAAAGTGTGCGCCGAGGGCTGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCAATCGGCTACCAAGTAAAGTGTGCGCCGAGGGCTGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGAGCATGACCAAGATCTGTGAAGCCCTTCCGCGCCCGCAACCC 1200
Db 1141 CCCAGCATCTTCCAGAGAGCATGACCAAGATCTGTGAAGCCCTTCCGCGCCCGCAACCC 1200
Qy 1201 GAGATGCTGATCTACCAAGTACATGACAGACTGTATGCTGGGCGACGACCTTGAGATTCG 1260
Db 1201 GAGATGCTGATCTACCAAGTACATGACAGACTGTATGCTGGGCGACGACCTTGAGATTCG 1260
Qy 1261 GAGCACCGGCGCAAGATCGAGAGAGCTGCGCAAGCACTGCTGGGCTTCAACCAAC 1320
Db 1261 GAGCACCGGCGCAAGATCGAGAGAGCTGCGCAAGCACTGCTGGGCTTCAACCAAC 1320
Qy 1321 CCCGCAAAAGACACCAAGAAAGAGCCCCCTTCTGTGTGATGGGCTTACGAGCTGCAACCC 1380
Db 1321 CCCGCAAAAGACACCAAGAAAGAGCCCCCTTCTGTGTGATGGGCTTACGAGCTGCAACCC 1380
Qy 1381 GACAAATGAGACGCTGACGCCATCGAGCTGCCGAGAAAGAGAGCTGACCCGTGAACGAC 1440
Db 1381 GACAAATGAGACGCTGACGCCATCGAGCTGCCGAGAAAGAGAGCTGACCCGTGAACGAC 1440
Qy 1441 ATTCAGAACTGTGTGGGCAAGCTGAACTGTGGGCGACGACAGATCTACCCCGCATCAAGTGT 1500
Db 1441 ATTCAGAACTGTGTGGGCAAGCTGAACTGTGGGCGACGACAGATCTACCCCGCATCAAGTGT 1500
Qy 1501 CGCCAGCTGTGCAAGCTGTGTGGGCGGCGCAAGGCCCTGACGACATCTGTGCCCCGTGAC 1560
Db 1501 CGCCAGCTGTGCAAGCTGTGTGGGCGGCGCAAGGCCCTGACGACATCTGTGCCCCGTGAC 1560
Qy 1561 GAGGAGGCGGAGCTGAGCTGAGCCGAGAAACCGGAGATCTGTGGGCGAGCCCGTGAACGAG 1620
Db 1561 GAGGAGGCGGAGCTGAGCTGAGCCGAGAAACCGGAGATCTGTGGGCGAGCCCGTGAACGAG 1620
Qy 1621 GTGTACTAGACCCCGAGCAAGGACTGTGTGGCGGAGATCGAAGAGAGGCGCACGACGAG 1680
Db 1621 GTGTACTAGACCCCGAGCAAGGACTGTGTGGCGGAGATCGAAGAGAGGCGCACGACGAG 1680
Qy 1681 TGGACTTACCAATCTAACGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGAAG 1740
Db 1681 TGGACTTACCAATCTAACGAGAGCCCTTCAAGAACTTGAAGACCGGCAAGTACGCGAAG 1740
Qy 1741 ATGCGCACGCGCCACGACCAAGAGCTGAAGAGCTGAACGAGGCGGTGAGAAAGATCGCC 1800
Db 1741 ATGCGCACGCGCCACGACCAAGAGCTGAAGAGCTGAACGAGGCGGTGAGAAAGATCGCC 1800
Qy 1801 ATGAGAGAGATCTGTGTGTGGGCGAAGACCCCGAAGTTCGCTGTGCCATTCAGAGAGAG 1860
Db 1801 ATGAGAGAGATCTGTGTGTGGGCGAAGACCCCGAAGTTCGCTGTGCCATTCAGAGAGAG 1860

Qy 1861 ACTTGGAGAGACTGTGTGACCGAATTAATGAGAGGCACTGTGATCCCGAGTGGAGATTTC 1920
Db 1861 ACTTGGAGAGACTGTGTGACCGAATTAATGAGAGGCACTGTGATCCCGAGTGGAGATTTC 1920
Qy 1921 GTGAACACCCCGCCCTGTGTGAGAGCTGTGTGTACAGCTGTGAGAAAGAGCCATCATCGAC 1980
Db 1921 GTGAACACCCCGCCCTGTGTGAGAGCTGTGTGTGTACAGCTGTGAGAAAGAGCCATCATCGAC 1980
Qy 1981 GCCGAGACCTTCTTACGTGTGACGCGCGCGCAACCCGAGAGCCAAATGTGGCAAGGCCGCG 2040
Db 1981 GCCGAGACCTTCTTACGTGTGACGCGCGCGCAACCCGAGAGCCAAATGTGGCAAGGCCGCG 2040
Qy 2041 TACGTGACCGAACCGGCGCGCGAGAAAGATCGTGAAGCTGTGACCGAGACCAAGCAAGAA 2100
Db 2041 TACGTGACCGAACCGGCGCGCGAGAAAGATCGTGAAGCTGTGACCGAGACCAAGCAAGAA 2100
Qy 2101 ACCGAGCTGACAGGACCATCGAGCTGCGCCCTGTGACAGACGCGAGCGAGGTGAACATCGT 2160
Db 2101 ACCGAGCTGACAGGACCATCGAGCTGCGCCCTGTGACAGACGCGAGCGAGGTGAACATCGT 2160
Qy 2161 ACCGAGCGCAGTACGCGCTGTGGCATCATCCAGGCCAGCCGACCAAGAGCGAGAGCGAG 2220
Db 2161 ACCGAGCGCAGTACGCGCTGTGGCATCATCCAGGCCAGCCGACCAAGAGCGAGAGCGAG 2220
Qy 2221 CTGTGTGAACCAAGATCATGAGAGAGCTGTCAAGAAAGAGAAAGTGTACTGTGAGTGGG 2280
Db 2221 CTGTGTGAACCAAGATCATGAGAGAGAGCTGTCAAGAAAGAGAAAGTGTACTGTGAGTGGG 2280
Qy 2281 CCCGCCCAAGAGGAGCATCGGCGGCAACGAGAGCATGCAAGCTGTGTGAGAGAGGAGCATC 2340
Db 2281 CCCGCCCAAGAGGAGCATCGGCGGCAACGAGAGCATGCAAGCTGTGTGAGAGAGGAGCATC 2340
Qy 2341 CGCAAGTGTGTCTTCTGTGACGCGCATCGATGCGGAGATCGATGCAAGCTGTGTGAGAG 2400
Db 2341 CGCAAGTGTGTCTTCTGTGACGCGCATCGATGCGGAGATCGATGCAAGCTGTGTGAGAG 2400
Qy 2401 GACCTGTACGTGTGGGAGCGGCGGCGCTTGAAGTCAATTAAGCTTCCGCGGCTGTGAC 2460
Db 2401 GACCTGTACGTGTGGGAGCGGCGGCGCTTGAAGTCAATTAAGCTTCCGCGGCTGTGAC 2460
Qy 2461 GGTGAATTC 2469
Db 2461 GGTGAATTC 2469

RESULT 3
ID ACA03548
ID ACA03548 standard; DNA; 2457 BP.
XX
AC ACA03548;
XX
DT 22-MAY-2003 (first entry)
XX
XX
DE Synthetic DNA encoding immunogenic HIV peptide #31.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; de.
OS Synthetic.
XX
PN MO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021421.
XX
PR 05-JUL-2001; 2001US-030192P.
PR 31-AUG-2001; 2001US-031686P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.

XX (CHTR) CHIRON CORP.
 PA Zur Megele J, Barnett SW, Lian Y;
 PI WPI; 2003-221602/21.
 DR
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C
 PT polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a subject,
 PT especially humans.
 XX
 PS Example 1; Fig 36; 262pp; English.
 XX
 XX The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for generating
 CC an immune response (e.g. a humoral or cellular immune response) in, a
 CC subject such as a mammal, particularly a human. This sequence encodes a
 CC human immunodeficiency virus immunogenic peptide
 XX
 SQ Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
 Query Match 99.5%; Score 2457; DB 8; Length 2457;
 Best Local Similarity 100.0%; Pred No.2.6e-296;
 Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 GCCACCATGCGCGAGCCATGAGCCAGGCCACCAAGCCCAATCTTGTAGAGCCGACG 66
 Db 1 GCCACCATGCGCGAGCCATGAGCCAGGCCACCAAGCCCAATCTTGTAGAGCCGACG 60
 QY 67 AACTTCAAGAGGCGCCCAAGGCGCATCATCAAGTGTCTTCAATCTGCGCAAGAGGCGCATC 126
 Db 61 AACTTCAAGAGGCGCCCAAGGCGCATCATCAAGTGTCTTCAATCTGCGCAAGAGGCGCATC 120
 QY 127 GCCCGCACTGCGCGCGCGCCCGCGCAAGAGGCGCTGTGAGAGTGGCGCAAGAGGCGCGAC 186
 Db 121 GCCCGCACTGCGCGCGCGCCCGCGCAAGAGGCGCTGTGAGAGTGGCGCAAGAGGCGCGAC 180
 QY 187 CAGATGAAGACTGCAACCGAGCGCAGGCCCACTTCTTCCGCGAGGAGACTTGGCTTCCG 246
 Db 181 CAGATGAAGACTGCAACCGAGCGCAGGCCCACTTCTTCCGCGAGGAGACTTGGCTTCCG 240
 QY 247 CAGGCGAAGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAAGCCCGCACGCGCG 306
 Db 241 CAGGCGAAGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCCCAAGCCCGCACGCGCG 300
 QY 307 GAGCTGCAAGTGGCGCGCGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
 Db 301 GAGCTGCAAGTGGCGCGCGCAACCCCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
 QY 367 CTGGAATCTTCCCGCAGATCAACCTGTGAGAGCGCCCTGTGTAGCATCAAGAGTGGCGCG 426
 Db 361 CTGGAATCTTCCCGCAGATCAACCTGTGTGAGAGCGCCCTGTGTAGCATCAAGAGTGGCGCG 420
 QY 427 CAGATCAAGAGGCGCTGTGTGACACCGGCGCGCAGCAGCACCGTGTCTGAGAGATGAGC 486
 Db 421 CAGATCAAGAGGCGCTGTGTGACACCGGCGCGCAGCAGCACCGTGTCTGAGAGATGAGC 480
 QY 487 CTGCGCGGCAAGTGGAGGCCCAAGATGATGCGGCGCATCGGCGCTTCAATCAAGTGGCGG 546
 Db 481 CTGCGCGGCAAGTGGAGGCCCAAGATGATGCGGCGCATCGGCGCTTCAATCAAGTGGCGG 540
 QY 547 CAGTACGACCAAGATCTGTATGAGATCTGCGCGCAAGAGGCGCATCGGCGCTGTGATC 606
 Db 541 CAGTACGACCAAGATCTGTATGAGATCTGCGCGCAAGAGGCGCATCGGCGCTGTGATC 600

QY 607 GGGCCCCACCCCGTGAACATCATCGGCGCAACATGTGACCCAGCTGGGCTGACCCCTG 666
 Db 601 GGGCCCCACCCCGTGAACATCATCGGCGCAACATGTGACCCAGCTGGGCTGACCCCTG 660
 QY 667 AACTTCCCATGAGCGCCCATCGAGACCGTGTCCCGTGAAGCTGAAGCCCGGCAATGACCGG 726
 Db 661 AACTTCCCATGAGCGCCCATCGAGACCGTGTCCCGTGAAGCTGAAGCCCGGCAATGACCGG 720
 QY 727 CCCAAGGTGAGCAAGTGGCGCTGTGACCGAGGAGGAAGATTAAGGCGCTTGACCCGCTATGCG 786
 Db 721 CCCAAGGTGAGCAAGTGGCGCTGTGACCGAGGAGGAAGATTAAGGCGCTTGACCCGCTATGCG 780
 QY 787 GAGGAGATGAGAAAGAGGCGCAAGATTCACAGATTCGCGCCGAGAACCCCTTACACACCC 846
 Db 781 GAGGAGATGAGAAAGAGGCGCAAGATTCACAGATTCGCGCCGAGAACCCCTTACACACCC 840
 QY 847 CCGGTGTTCCCATGAGAAAGAGGAGCAGACCAAGTGGCGCAAGCTGTGATCTTCCGCG 906
 Db 841 CCGGTGTTCCCATGAGAAAGAGGAGCAGACCAAGTGGCGCAAGCTGTGATCTTCCGCG 900
 QY 907 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGACAGCTGGGCAATCCCCACCCCGCG 966
 Db 901 GAGCTGAACAGCGCACCCAGGACTTCTGGAGGTGACAGCTGGGCAATCCCCACCCCGCG 960
 QY 967 GGCCTGAAGAAAGAGAGCGGTGACCGTGTGAGCGTGGGCGAGCGCTTCAAGCGGTG 1026
 Db 961 GGCCTGAAGAAAGAGAGCGGTGACCGTGTGAGCGTGGGCGAGCGCTTCAAGCGGTG 1020
 QY 1027 CCCCTGAGACGAGGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAG 1086
 Db 1021 CCCCTGAGACGAGGACTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAG 1080
 QY 1087 ACCCGCGGATCGGCTTCAAGTACAACTGTGCTGCCAGGCGCTGAGAGGCGAGCCCGACG 1146
 Db 1081 ACCCGCGGATCGGCTTCAAGTACAACTGTGCTGCCAGGCGCTGAGAGGCGAGCCCGACG 1140
 QY 1147 ATCTTCCAGAGAGCATGACCAAGATCTCTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
 Db 1141 ATCTTCCAGAGAGCATGACCAAGATCTCTGAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
 QY 1207 GTGATCTACAGTACATGAGACGACTGTAGTGTGGGAGCGACCTTGAGATTCGGCGAGCAC 1266
 Db 1201 GTGATCTACAGTACATGAGACGACTGTAGTGTGGGAGCGACCTTGAGATTCGGCGAGCAC 1260
 QY 1267 CGCGCCAGATGAGAGAGCTGTGCGCAAGCACCTGTGCGTGGGCGCTTCAACACCCCGAC 1326
 Db 1261 CGCGCCAGATGAGAGAGCTGTGCGCAAGCACCTGTGCGTGGGCGCTTCAACACCCCGAC 1320
 QY 1327 AAGAAACACCAAGAGAGGCGCCCTTCTGTGTGATGTGGCTTACAGGCTGACACCCCGCAAG 1386
 Db 1321 AAGAAACACCAAGAGAGGCGCCCTTCTGTGTGATGTGGCTTACAGGCTGACACCCCGCAAG 1380
 QY 1387 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGCTGAGACGCTGAACGATCCAG 1446
 Db 1381 TGGACCGTGCAGCCCATCGAGCTGCGCGAGAGAGAGCTGAGACGCTGAACGATCCAG 1440
 QY 1447 AAGCTGTGTGGCAAGCTGAATCTGTGGCGAGCAGATCTTACCCCGCATCAAGTGGCGCGAG 1506
 Db 1441 AAGCTGTGTGGCAAGCTGAATCTGTGGCGAGCAGATCTTACCCCGCATCAAGTGGCGCGAG 1500
 QY 1507 CTGTGCAAGCTGTGTGCGCGCGCGCAAGGCGCTTGAACCAATGTGCGCTTGAACCGAGAG 1566
 Db 1501 CTGTGCAAGCTGTGTGCGCGCGCGCAAGGCGCTTGAACCAATGTGCGCTTGAACCGAGAG 1560
 QY 1567 GCGGAGCTGAGACTGCGCGAGAACCGCGAGATCTTGTGCGAGGCGCGTGTGACCGGCGGTAC 1626
 Db 1561 GCGGAGCTGAGACTGCGCGAGAACCGCGAGATCTTGTGCGAGGCGCGTGTGACCGGCGGTAC 1620
 QY 1627 TACGACCCGAGCAAGGACTGTGTGCGAGATTCAGAAACAGGCGCAACCAAGTGTGAC 1686
 Db 1621 TACGACCCGAGCAAGGACTGTGTGCGAGATTCAGAAACAGGCGCAACCAAGTGTGAC 1680
 QY 1687 TACGAGATCTACAGAGAGCCCTTCAAGAACTGTGAAGACCGGCGCAAGTACCCCAAGATGCCG 1746

```

Db 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACTGAAAGACCGGAGAGTACGCAAGATGCGC 1740
Qy 1747 ACCGCCCAACCAAGACGTTGAAGACGCTGACCGAGGCGCTGACAGAGATCCCATGAG 1806
Db 1741 ACCGCCCAACCAAGACGTTGAAGACGCTGACCGAGGCGCTGACAGAGATCCCATGAG 1800
Qy 1807 AGCATGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGCTGG 1866
Db 1801 AGCATGTATCTGGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGCTGG 1860
Qy 1867 GAGACTGTGTGAGACCGACTACTGTGAGAGCGCACCTGATCCCGAGTGGAGTTCTGTAAC 1926
Db 1861 GAGACTGTGTGAGACCGACTACTGTGAGAGCGCACCTGATCCCGAGTGGAGTTCTGTAAC 1920
Qy 1927 ACCCCCCCTTGTGTGAAGTGTGTGTACCAAGCTGAGAGAGAGCCCATCATCCGCGCGAG 1986
Db 1921 ACCCCCCCTTGTGTGAAGTGTGTGTACCAAGCTGAGAGAGAGCCCATCATCCGCGCGAG 1980
Qy 1987 ACCTTCTAGTGAACGGGCGCGCAACCGGAGAACCAAGATGGGCAAGGCGGCTTACGTT 2046
Db 1981 ACCTTCTAGTGAACGGGCGCGCAACCGGAGAACCAAGATGGGCAAGGCGGCTTACGTT 2040
Qy 2047 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGTACCGAGACCAACCAAGAGACCGAG 2106
Db 2041 ACCGACCGGCGCGCGAGAGATCTGTGAGCTGTACCGAGACCAACCAAGAGACCGAG 2100
Qy 2107 CTGCAAGGCGCATCAAGTGTGCGCTTGTGAGAGACAGCGGAGAGAGTGAACATCTGTACCGAC 2166
Db 2101 CTGCAAGGCGCATCAAGTGTGCGCTTGTGAGAGACAGCGGAGAGAGTGAACATCTGTACCGAC 2160
Qy 2167 AGCCAGTACGCGCTGTGAGATCTCAAGCGCCAGCCGCAAGAGAGAGAGAGAGCTGGT 2226
Db 2161 AGCCAGTACGCGCTGTGAGATCTCAAGCGCCAGCCGCAAGAGAGAGAGAGAGCTGGT 2220
Qy 2227 AACGAGATCATGAGAGAGTGTATCAAGAGAGAGAGTGTATCTGAGTGGTGCCTGCGC 2286
Db 2221 AACGAGATCATGAGAGAGTGTATCAAGAGAGAGAGTGTATCTGAGTGGTGCCTGCGC 2280
Qy 2287 CACAAGGGGCGATCGCGGCAACGACGATCGACAGAGCTGTGTGACAGAGGCGATCCGCAAG 2346
Db 2281 CACAAGGGGCGATCGCGGCAACGACGATCGACAGAGCTGTGTGACAGAGGCGATCCGCAAG 2340
Qy 2347 GGTGCTGTTCTGGAAGCGGATCGATGGGCGGCGATCTGATCTACAGAGATGAGACGCTG 2406
Db 2341 GGTGCTGTTCTGGAAGCGGATCGATGGGCGGCGATCTGATCTACAGAGATGAGACGCTG 2400
Qy 2407 TACGTGGGAGCGCGCGCTTGAAGATCGATTAAAGCTTCCCGGGCTTACGACCGGT 2463
Db 2401 TACGTGGGAGCGCGCGCTTGAAGATCGATTAAAGCTTCCCGGGCTTACGACCGGT 2457

```

RESULT 4

ADCl3266
ID ADCl3266 standard; DNA; 2457 BP.

ADCl3266;

18-DEC-2003 (first entry)

DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.

expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; Prot;
Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.

Human immunodeficiency virus.

WO2003004620-A2.

16-JAN-2003.

05-JUL-2002; 2002WO-US021420.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```

PR 05-JUL-2001; 2001US-030192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STERILEBOSCH.
PI Zur Megele U, Barnette SW, Lian Y, Engelbrecht S, Van Rensburg BJ,
DR WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 42; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
XX
XX Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;

```

Query Match 99.5%; Score 2457; DB 10; Length 2457;

Best Local Similarity 100.0%; Pred. No. 2.66-286; Mismatches 0; Indels 0; Gaps 0;

Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 7 GCCACCATGCGCGAGGCGCATGAGCGGCGACCGGCGCAACATCTTGTGAGCGGAGC 66
Db 1 GCCACCATGCGCGAGGCGCATGAGCGGCGACCGGCGCAACATCTTGTGAGCGGAGC 60
Qy 67 AACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGGCAAGAGGCGCATC 126
Db 61 AACTTCAAGGGGCCCAAGCGCATCATCAAGTGTCTTCAACTGTGGGCAAGAGGCGCATC 120
Qy 127 GCCCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 186
Db 121 GCCCGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 180
Qy 187 CAGATGAAGAGACTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 246
Db 181 CAGATGAAGAGACTGACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 240
Qy 247 CAGGCGAAGCGCGCGAGTTCGCCAGCGAGCAAGAACCGGCGCAAGCGCCGCGCGC 306
Db 241 CAGGCGAAGCGCGCGAGTTCGCCAGCGAGCAAGAACCGGCGCAAGCGCCGCGCGC 300
Qy 307 GAGCTGCAAGTGTGCGGCGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 366
Db 301 GAGCTGCAAGTGTGCGGCGGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC 360
Qy 367 CTGAATCTTCCCGCAGATCACTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 426
Db 361 CTGAATCTTCCCGCAGATCACTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGC 420
Qy 427 CAGATCAAGAGAGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGC 486
Db 421 CAGATCAAGAGAGCGCGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGC 480
Qy 487 CTGCGCGGCAAGTGAAGACCGCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGC 546
Db 481 CTGCGCGGCAAGTGAAGACCGCAAGATGATCGCGCGCGCGCGCGCGCGCGCGCGCGC 540
Qy 547 CAGTACGACCAAGTCTGTATGAGATCTGCGGCGCAAGAGGCGCATCGGCGCGCGTGC 606

```


XX 17-JAN-2002.
PD 05-JUL-2001; 2001MO-US021241.
XX 05-JUL-2001; 2000US-00610313.
XX 05-JUL-2001; 2000US-00610313.
XX (CHIR) CHIRON CORP.
XX (UYST-) UNIT STELLENBOSCH.
XX Zur Megele J, Barnett SW, Engelbrecht S, Van Rensburg EJ,
XX WPI; 2002-154920/20.
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
XX applications including DNA immunization or generation of packaging cell
XX lines, particularly in gene therapy.
XX Claim 1; Fig 9; 233pp; English.
XX The present invention describes expression cassettes comprising a
XX polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
XX type C polypeptides. The expression cassettes comprise any of the HIV
XX type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
XX (I). (i) have immunostimulant activity and can be used in gene therapy.
XX The HIV type C polynucleotides are useful in applications including DNA
XX immunisation, generation of packaging cell lines, and production of HIV
XX Type C proteins. The polynucleotides are particularly useful in gene
XX therapy and DNA immunisation applications. ABL3942 to ABL40054 and
XX ABL06204 to ABL06215 represent sequences used in the exemplification of
XX the present invention
XX
XX Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 98.9%; Score 2442.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 1.7e-294;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GTGCAAGCCAGCATGCGAGGCGCATGAGCGCAAGGCGCAATCTGATGCGAG 60
DB 1 GTGCAAGCCAGCATGCGAGGCGCATGAGCGCAAGGCGCAATCTGATGCGAG 60
QY 61 CGCAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGC 120
DB 61 CGCAGCACTTCAAGGCGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGC 120
QY 121 CACATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGCGTCTGGAAGTGCAGAGAG 180
DB 121 CACATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGCGTCTGGAAGTGCAGAGAG 180
QY 181 GGCACACAGATGAAGGACTGCAACCGAGCGCGCAAGCTTCTCCGAGAGGACTGCGGC 240
DB 181 GGCACACAGATGAAGGACTGCAACCGAGCGCGCAAGCTTCTCCGAGAGGACTGCGGC 240
QY 241 TTCCGCCAGAGGCAAGGCGCCGCGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCGACC 300
DB 241 TTCCGCCAGAGGCAAGGCGCCGCGAGTTCCCGAGGAGCAGAACCGCGCAACAGCCGACC 300
QY 301 AGCCGCGAGCTGAGGTCGCGCGGCAACACCCCGCAGAGGCGCGCGCGCGAGCGCGAG 360
DB 301 AGCCGCGAGCTGAGGTCGCGCGGCAACACCCCGCAGAGGCGCGCGCGCGAGCGCGAG 360
QY 361 GGCACCTGAACTTCCCGCAGATCAACCTGTGCGAGCGCCCTGTGTTGAGCATCAAGTG 420
DB 361 GGCACCTGAACTTCCCGCAGATCAACCTGTGCGAGCGCCCTGTGTGAGCATCAAGTG 420
QY 421 GCGCGCAGATCAAGAGGCGCCGTGCTGACACCGCGCGCGAGCAGACCGTGTCTGAGAG 480
DB 421 GCGCGCAGATCAAGAGGCGCCGTGCTGACACCGCGCGCGAGCAGACCGTGTCTGAGAG 480
QY 481 ATGAGCTTGCCTGCGGCAAGTGAAGGCCCAAGATGATGCGCGGATCGGCGGCTTCATCAAG 540
DB 481 ATGAGCTTGCCTGCGGCAAGTGAAGGCCCAAGATGATGCGCGGATCGGCGGCTTCATCAAG 540

QY 541 GTGCGCAGTACGACCAAGATCTGATGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600
DB 541 GTGCGCAGTACGACCAAGATCTGATGAGATCTGCGGCAAGAGGCCATCGGCAACCGTG 600
QY 601 CTGATCGGCGCCCAAGCGCGTGAACATCATCGGCGCGCAACATGCTGAGCCGCTGCGC 660
DB 601 CTGATCGGCGCCCAAGCGCGTGAACATCATCGGCGCGCAACATGCTGAGCCGCTGCGC 660
QY 661 ACCCTGAACCTTCCCATATGAGCCCATGAGACCGTGCCTGTGAACCTGAAGCCCGGCA 720
DB 661 ACCCTGAACCTTCCCATATGAGCCCATGAGACCGTGCCTGTGAACCTGAAGCCCGGCA 720
QY 721 GACGCGCCCAAGGAGTGAAGAGTGCCTGACCCGAGAGAAATCAAGGCGCTGACCGGC 780
DB 721 GACGCGCCCAAGGAGTGAAGAGTGCCTGACCCGAGAGAAATCAAGGCGCTGACCGGC 780
QY 781 ATTCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATCGGCGCGAGAAACCCCTAC 840
DB 781 ATTCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATCGGCGCGAGAAACCCCTAC 840
QY 841 AACACCCCGCTGTTCCTTCAAGAAAGAGACAGACCAAGTGCAGAGCTGTGTGAGC 900
DB 841 AACACCCCGCTGTTCCTTCAAGAAAGAGACAGACCAAGTGCAGAGCTGTGTGAGC 900
QY 901 TTCGCGAGCTGAACAAGCGCACCCGAGACTTCTGAGAGGTGACAGCTGAGGCAATCCGCC 960
DB 901 TTCGCGAGCTGAACAAGCGCACCCGAGACTTCTGAGAGGTGACAGCTGAGGCAATCCGCC 960
QY 961 CCGCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGAGTGAAGCGGCGCAACCTTC 1020
DB 961 CCGCGCGCGCTGAAGAAAGAGAGCGTGAACCGTGTGAGAGTGAAGCGGCGCAACCTTC 1020
QY 1021 AGCGTGCCTGTGAGAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGACATCAAC 1080
DB 1021 AGCGTGCCTGTGAGAGAGACTTCCGCAAGTACACCGCTTCAACATCCCGACATCAAC 1080
QY 1081 AACGAGACCCCGGAGATCGCTACCAAGTACAGAGTGTGCGCCGAGGCTGAGAGGAGC 1140
DB 1081 AACGAGACCCCGGAGATCGCTACCAAGTACAGAGTGTGCGCCGAGGCTGAGAGGAGC 1140
QY 1141 CCGACATCTTCAAGAGCAGATGAACCAAGTCTGAGGCGCTTCGCGCGCGCAACCGC 1200
DB 1141 CCGACATCTTCAAGAGCAGATGAACCAAGTCTGAGGCGCTTCGCGCGCGCAACCGC 1200
QY 1201 GAGATCGATTTACAGATGACAGACCTGTACGTTGAGGAGGAGCCTTGAAGTTCGCG 1260
DB 1201 GAGATCGATTTACAGATGACAGACCTGTACGTTGAGGAGGAGCCTTGAAGTTCGCG 1260
QY 1261 CAGCAACCGCGCAAGTGAAGGAGCTGCGCAAGCCTGTCGCGCGGCGCTTCAACCGC 1320
DB 1261 CAGCAACCGCGCAAGTGAAGGAGCTGCGCAAGCCTGTCGCGCGGCGCTTCAACCGC 1320
QY 1321 CCGCAAGAAAGCAACAGAGAGGCGCCCTTCTGTGATGAGCTACAGAGTGAACCGC 1380
DB 1321 CCGCAAGAAAGCAACAGAGAGGCGCCCTTCTGTGATGAGCTACAGAGTGAACCGC 1380
QY 1315 CCGCAAGAAAGCAACAGAGAGGCGCCCTTCTGTGATGAGCTACAGAGTGAACCGC 1374
DB 1315 CCGCAAGAAAGCAACAGAGAGGCGCCCTTCTGTGATGAGCTACAGAGTGAACCGC 1374
QY 1381 GACAAAGTGAACCTGACAGCCATGAGCTGCTCGAGAAAGAGAGCTGACCTGTAAACGAC 1440
DB 1381 GACAAAGTGAACCTGACAGCCATGAGCTGCTCGAGAAAGAGAGCTGACCTGTAAACGAC 1440
QY 1375 GACAAAGTGAACCTGACAGCCATGAGCTGCTCGAGAAAGAGAGCTGACCTGTAAACGAC 1434
DB 1375 GACAAAGTGAACCTGACAGCCATGAGCTGCTCGAGAAAGAGAGCTGACCTGTAAACGAC 1434
QY 1441 ATTCAGAGAGCTGTGAGGAGAGTGAACCTGAGCAGCAGATTAACCCGCGGCTCAAGAG 1500
DB 1441 ATTCAGAGAGCTGTGAGGAGAGTGAACCTGAGCAGCAGATTAACCCGCGGCTCAAGAG 1500
QY 1435 ATTCAGAGAGCTGTGAGGAGAGTGAACCTGAGCAGCAGATTAACCCGCGGCTCAAGAG 1494
DB 1435 ATTCAGAGAGCTGTGAGGAGAGTGAACCTGAGCAGCAGATTAACCCGCGGCTCAAGAG 1494
QY 1501 CCGCAAGTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGCAATCGTGCCTGAC 1560
DB 1501 CCGCAAGTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGCAATCGTGCCTGAC 1560
QY 1495 CCGCAAGTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGCAATCGTGCCTGAC 1554
DB 1495 CCGCAAGTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGCAATCGTGCCTGAC 1554
QY 1561 GAGAGGCGGAGCTGAGAGTGCAGAAACCGCGAGATCTGCGGAGGCGCTGCAACGCGC 1620
DB 1561 GAGAGGCGGAGCTGAGAGTGCAGAAACCGCGAGATCTGCGGAGGCGCTGCAACGCGC 1620
QY 1555 GAGAGGCGGAGCTGAGAGTGCAGAAACCGCGAGATCTGCGGAGGCGCTGCAACGCGC 1614
DB 1555 GAGAGGCGGAGCTGAGAGTGCAGAAACCGCGAGATCTGCGGAGGCGCTGCAACGCGC 1614

D 361 GGCACCTTGAACTTCCCGCAGATCACTCTGTGCAAGCCCTCTGTGATCAAGGTG 420
Q 421 GGGCGGCAGATCAAGAGGCGCTGTGCAACCGGCGCGCAAGACACCGTGTGAGAG 480
D 421 GGGCGGCAGATCAAGAGGCGCTGTGCAACCGGCGCGCAAGACACCGTGTGAGAG 480
Q 481 ATGAGCTTCCCGCAAGTGAAGCCCAAGATATCGCGGATCGCGGCTTCAACAAG 540
D 481 ATGAGCTTCCCGCAAGTGAAGCCCAAGATATCGCGGATCGCGGCTTCAACAAG 540
Q 541 GTGCGCAGATCAAGACAGATCTGTATCGAGATCTGCGGCAAGAGCCATCGGACCTG 600
D 541 GTGCGCAGATCAAGACAGATCTGTATCGAGATCTGCGGCAAGAGCCATCGGACCTG 600
Q 601 CTGATCGGCCCCACCGCGTGAACATCATCGGCGGCAATGCTGACCCAGCTGGGCTGC 660
D 601 CTGATCGGCCCCACCGCGTGAACATCATCGGCGGCAATGCTGACCCAGCTGGGCTGC 660
Q 661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 720
D 661 ACCCTGAACCTTCCCATCAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 720
Q 721 GACGCGCCCCAGAGTGAAGAGTGGCCCTGACCCGAGAGAGATCAAGGCTTGAACCGCC 780
D 721 GACGCGCCCCAGAGTGAAGAGTGGCCCTGACCCGAGAGAGATCAAGGCTTGAACCGCC 780
Q 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGAGAACCCCTAC 840
D 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGAGAACCCCTAC 840
Q 841 AACACCCCGTGTTCGCATCAAGAGAGAGACAGACCAAGTGGCGGACGTGTGAGAC 900
D 841 AACACCCCGTGTTCGCATCAAGAGAGAGACAGACCAAGTGGCGGACGTGTGAGAC 900
Q 901 TTCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGTGAGAGTGAAGTGGGATCCGCCAC 960
D 901 TTCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGTGAGAGTGAAGTGGGATCCGCCAC 960
Q 961 CCGCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGACCGCTACCTTC 1020
D 961 CCGCGCGGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGACCGCTACCTTC 1020
Q 1021 AGCGTGCCTTGAAGAGAGACTTTCGCAAGTACACCGGCTTCAACATCCCGACAGATCAAC 1080
D 1021 AGCGTGCCTTGAAGAGAGACTTTCGCAAGTACACCGGCTTCAACATCCCGACAGATCAAC 1080
Q 1081 AAGGAGACCCCGGCAATCCGCTACCAAGTCAAGTGTGCTGCCAGGCGTGAAGGCGACG 1140
D 1081 AAGGAGACCCCGGCAATCCGCTACCAAGTCAAGTGTGCTGCCAGGCGTGAAGGCGACG 1140
Q 1141 CCAGCATCTTCAAGAGAGAGATCAAGATCTGTGAGCGCTTCGCGCGCGCAACCCG 1200
D 1141 CCAGCATCTTCAAGAGAGAGATCAAGATCTGTGAGCGCTTCGCGCGCGCAACCCG 1200
Q 1201 GAGATCTGTATCTACAGTACATGACAGACTGTATGCTGTGAGAGAGACTTGTGAGATCGAG 1260
D 1201 GAGATCTGTATCTACAGTACATGACAGACTGTATGCTGTGAGAGAGACTTGTGAGATCGAG 1260
Q 1261 CAGCACCGGCGCAAGTCAAGAGAGTGTGCGCAAGCACCTGTGCGGCTTCAACACG 1320
D 1261 CAGCACCGGCGCAAGTCAAGAGAGTGTGCGCAAGCACCTGTGCGGCTTCAACACG 1320
Q 1321 CCGGCAAGAGAGACCAAGAGAGCGGCTTCTGTGTGATGCGGCTGAAGCTGACCGCC 1380
D 1321 CCGGCAAGAGAGACCAAGAGAGCGGCTTCTGTGTGATGCGGCTGAAGCTGACCGCC 1380
Q 1381 GACAGTGAAGCTGTGAGAGCTTCAAGAGAGTGTGCGGAGAGAGAGAGAGCTGACCGCTGAACGAC 1440
D 1381 GACAGTGAAGCTGTGAGAGCTTCAAGAGAGTGTGCGGAGAGAGAGAGAGCTGACCGCTGAACGAC 1440
Q 1441 ATTCAGAGAGCTGTGTGAGAGAGCTGAACTGTGGCCAGGCAAGCTTACCCCGGATCAAGGTG 1500
D 1441 ATTCAGAGAGCTGTGTGAGAGAGCTGAACTGTGGCCAGGCAAGCTTACCCCGGATCAAGGTG 1500
Q 1501 CGCCAGCTGTGAAGCTGTGTGCGCGGCAAGAGCCCTGACCGACATGTGTGCCCTGAC 1560
D 1501 CGCCAGCTGTGAAGCTGTGTGCGCGGCAAGAGCCCTGACCGACATGTGTGCCCTGAC 1560
Q 1561 GAGAGAGCGAGCTGTGAAGCTGTGCGGAGAACCGGAGATCTGTGCGGAGCCGTGCAAG 1620
D 1561 GAGAGAGCGAGCTGTGAAGCTGTGCGGAGAACCGGAGATCTGTGCGGAGCCGTGCAAG 1620
Q 1621 GTGTACTACAGACCCAGAGAGAGCTGTGTGCGGAGATCTGAGAGAGAGAGAGAGAG 1680
D 1621 GTGTACTACAGACCCAGAGAGAGCTGTGTGCGGAGATCTGAGAGAGAGAGAGAGAG 1680
Q 1681 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGCAAGTACGCAAG 1740
D 1681 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGCAAGTACGCAAG 1740
Q 1741 ATGCGCACCGCCCAACCAACAGCTGAAAGCACTGAACCGAGCGGTGAGAGATCGCC 1800
D 1741 ATGCGCACCGCCCAACCAACAGCTGAAAGCACTGAACCGAGCGGTGAGAGATCGCC 1800
Q 1801 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
D 1801 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Q 1861 ACCTGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
D 1861 ACCTGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Q 1921 GTGAACACCCCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
D 1921 GTGAACACCCCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Q 1981 GCGGAGAGCTTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
D 1981 GCGGAGAGCTTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Q 2041 TACGTGAG 2100
D 2041 TACGTGAG 2100
Q 2101 ACCGAGCTGAG 2160
D 2101 ACCGAGCTGAG 2160
Q 2161 ACCGAG 2220
D 2161 ACCGAG 2220
Q 2221 CTGTGAG 2280
D 2221 CTGTGAG 2280
Q 2281 CCGGCGCAAG 2340
D 2281 CCGGCGCAAG 2340
Q 2341 CGCAAG 2400
D 2341 CGCAAG 2400
Q 2401 GACCTGTAG 2460
D 2401 GACCTGTAG 2460
Q 2461 GGTGAATTC 2469
D 2461 GGTGAATTC 2469
Q 2469 GGTGAATTC 2463
D 2469 GGTGAATTC 2463

Q 1501 CGCCAGCTGTGAAGCTGTGTGCGCGGCAAGAGCCCTGACCGACATGTGTGCCCTGAC 1560
D 1501 CGCCAGCTGTGAAGCTGTGTGCGCGGCAAGAGCCCTGACCGACATGTGTGCCCTGAC 1560
Q 1561 GAGAGAGCGAGCTGTGAAGCTGTGCGGAGAACCGGAGATCTGTGCGGAGCCGTGCAAG 1620
D 1561 GAGAGAGCGAGCTGTGAAGCTGTGCGGAGAACCGGAGATCTGTGCGGAGCCGTGCAAG 1620
Q 1621 GTGTACTACAGACCCAGAGAGAGCTGTGTGCGGAGATCTGAGAGAGAGAGAGAGAG 1680
D 1621 GTGTACTACAGACCCAGAGAGAGCTGTGTGCGGAGATCTGAGAGAGAGAGAGAGAG 1680
Q 1681 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGCAAGTACGCAAG 1740
D 1681 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGCAAGTACGCAAG 1740
Q 1741 ATGCGCACCGCCCAACCAACAGCTGAAAGCACTGAACCGAGCGGTGAGAGATCGCC 1800
D 1741 ATGCGCACCGCCCAACCAACAGCTGAAAGCACTGAACCGAGCGGTGAGAGATCGCC 1800
Q 1801 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
D 1801 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Q 1861 ACCTGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
D 1861 ACCTGGAGAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Q 1921 GTGAACACCCCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
D 1921 GTGAACACCCCGCTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Q 1981 GCGGAGAGCTTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
D 1981 GCGGAGAGCTTCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Q 2041 TACGTGAG 2100
D 2041 TACGTGAG 2100
Q 2101 ACCGAGCTGAG 2160
D 2101 ACCGAGCTGAG 2160
Q 2161 ACCGAG 2220
D 2161 ACCGAG 2220
Q 2221 CTGTGAG 2280
D 2221 CTGTGAG 2280
Q 2281 CCGGCGCAAG 2340
D 2281 CCGGCGCAAG 2340
Q 2341 CGCAAG 2400
D 2341 CGCAAG 2400
Q 2401 GACCTGTAG 2460
D 2401 GACCTGTAG 2460
Q 2461 GGTGAATTC 2469
D 2461 GGTGAATTC 2469
Q 2469 GGTGAATTC 2463
D 2469 GGTGAATTC 2463

RESULT 7
ACN03547
ID ACN03547 standard; DNA; 2457 BP.

XX ACA03547;
 XX
 XX 22-MAY-2003 (first entry)
 XX
 XX Synthetic DNA encoding immunogenic HIV peptide #30.
 XX
 XX Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
 XX gene therapy; packaging cell line; humoral immune response;
 XX cellular immune response; gene delivery vector; DNA immunisation; da.
 XX
 OS Synthetic.
 PN WO2003004657-A1.
 PD 16-JAN-2003.
 XX
 XX 05-JUL-2002; 2002WO-US021421.
 PF
 XX 05-JUL-2001; 2001US-0303192P.
 PR 31-AUG-2001; 2001US-0316860P.
 PR 16-JAN-2002; 2002US-0349728P.
 PR 16-JAN-2002; 2002US-0349793P.
 PR 16-JAN-2002; 2002US-0349871P.
 XX
 PA (CHIR) CHIRON CORP.
 PI Zur Megede J, Barnett SW, Llan Y;
 XX MPI; 2003-221602/21.
 DR
 XX New synthetic polynucleotides encoding antigenic HIV type B and/or type C
 PT polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a subject,
 PT especially humans.
 PS
 XX Example 1; Fig 35; 262pp; English.
 XX
 CC The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for generating
 CC an immune response (e.g. a humoral or cellular immune response) in, a
 CC subject such as a mammal, particularly a human. This sequence encodes a
 CC human immunodeficiency virus immunogenic peptide
 XX
 SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;
 Query Match 98.6%; Score 2434.6; DB 8; Length 2457;
 Best Local Similarity 99.6%; Pred. No. 1.5e-293;
 Matches 2453; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 241 TTCCCCCAAGGGGCAAGCCCGGAGATTCCCAAGAGCAAGAACCGGGCCCAAGCCCCACC 300
 |||
 DB 241 TTCCCCCAAGGGGCAAGCCCGGAGATTCCCAAGAGCAAGAACCGGGCCCAAGCCCCACC 300
 QY 301 AGCCGCGAGCTTCAGGTGCGCGCGCAACACCCCGCAGAGGCGCGCCGAGCGCCAG 360
 |||
 DB 301 AGCCGCGAGCTTCAGGTGCGCGCGCAACACCCCGCAGAGGCGCGCCGAGCGCCAG 360
 QY 361 GGCACCTTGAACTTCCCCCAAGTCAACCTGTGGCAGCGCCCTTGTGAGCATCAAGGTG 420
 |||
 DB 361 GGCACCTTGAACTTCCCCCAAGTCAACCTGTGGCAGCGCCCTTGTGAGCATCAAGGTG 420
 QY 421 GGGCGCGAATCAAGAGGCGCTGTGACAACCGGGCGCCAGCAACCGTGTGAGGAG 480
 |||
 DB 421 GGGCGCGAATCAAGAGGCGCTGTGACAACCGGGCGCCAGCAACCGTGTGAGGAG 480
 QY 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCAAG 540
 |||
 DB 481 ATGAGCCTGCGCGCAAGTGAAGCCCAAGATGATCGCGGATCGCGGCTTCATCAAG 540
 QY 541 GTGCGCAATACCAACCAAGTCTGATCGAGTCTGCGCAAGAGGCCATCGGCACGTG 600
 |||
 DB 541 GTGCGCAATACCAACCAAGTCTGATCGAGTCTGCGCAAGAGGCCATCGGCACGTG 600
 QY 541 GTGCGCAATACCAACCAAGTCTGATCGAGTCTGCGCAAGAGGCCATCGGCACGTG 600
 |||
 DB 601 CTGATCGGCCCCAACCCTGTGAACATCATCGGCCCGCAATGCTGACCCAGCTGGGCTGC 660
 |||
 DB 601 CTGATCGGCCCCAACCCTGTGAACATCATCGGCCCGCAATGCTGACCCAGCTGGGCTGC 660
 QY 661 ACCCTGAATTTCCCATGAGCCCATGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
 |||
 DB 661 ACCCTGAATTTCCCATGAGCCCATGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
 QY 721 GACGCGCCCAAGGTGAAGGATGCGCCCTGACCGAGAGAAATCAAGCCCTGACCGCC 780
 |||
 DB 721 GACGCGCCCAAGGTGAAGGATGCGCCCTGACCGAGAGAAATCAAGCCCTGACCGCC 780
 QY 781 ATCTGGAAGATGGAAGAGGCGCAAGATCAACCAATGCGCCCGAGAACCCCTAC 840
 |||
 DB 781 ATCTGGAAGATGGAAGAGGCGCAAGATCAACCAATGCGCCCGAGAACCCCTAC 840
 QY 841 AACACCCCGTGTTCGCAATCAAGAGAGAGCAAGCAACCAAGTGGCGCAAGCTGTGAG 900
 |||
 DB 841 AACACCCCGTGTTCGCAATCAAGAGAGAGCAAGCAACCAAGTGGCGCAAGCTGTGAG 900
 QY 901 TTCCGGAAGTGAACAGGCGAACCGAGACTTCTGGAAGGTGAGCTGGCATCCCGAC 960
 |||
 DB 901 TTCCGGAAGTGAACAGGCGAACCGAGACTTCTGGAAGGTGAGCTGGCATCCCGAC 960
 QY 961 CCGCGCGGCTTGAAG 1020
 |||
 DB 961 CCGCGCGGCTTGAAG 1020
 QY 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCACATCCCGACATCAAC 1080
 |||
 DB 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCACATCCCGACATCAAC 1080
 QY 1081 AACGAGACCCCGGCAATCCGCTACAGTACCAAGTCTGCGCCGAGGCTTGAAGAGGAGC 1140
 |||
 DB 1081 AACGAGACCCCGGCAATCCGCTACAGTACCAAGTCTGCGCCGAGGCTTGAAGAGGAGC 1140
 QY 1141 CCGAGCATCTTCAAGAGAGAGATGACCAAGATCTTGAAGCTTCCGCGCCCGCAACCCC 1200
 |||
 DB 1141 CCGAGCATCTTCAAGAGAGAGATGACCAAGATCTTGAAGCTTCCGCGCCCGCAACCCC 1200
 QY 1201 GAGATGATCTCAAGAGTACATGACGACCTGTAGTGGGCGAGACCTTGAGATCGG 1260
 |||
 DB 1201 GAGATGATCTCAAGAGTACATGACGACCTGTAGTGGGCGAGACCTTGAGATCGG 1260
 QY 1261 CAGCACCGGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCGCTTCAACACC 1320
 |||
 DB 1261 CAGCACCGGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCGCTTCAACACC 1320
 QY 1255 CAGCACCGGCGCAAGATCGAGAGCTGCGCAAGCACTGTGCGCTTCAACACC 1314
 |||
 DB 1321 CCGCAAGAGAGCAACAG 1380

Db 121 CACATGCGCCGCACTGCGCGCCGCCCGCCGAGAGAGGCTGCGAGAGTGCAGAGAG 180
Qy 181 GGCACACAGATGAGAGACTGCAACGAGCGCCAGGCCCACTTCTTCCGACGAGACTGAGCC 240
Db 181 GGCACACAGATGAGAGACTGCAACGAGCGCCAGGCCCACTTCTTCCGACGAGACTGAGCC 240
Qy 241 TTCCCCCAAGGCGAGAGGCGCGGAGTTCCCAAGGAGAGAGACCGGCCCAAGGCCCAAC 300
Db 241 TTCCCCCAAGGCGAGAGGCGCGGAGTTCCCAAGGAGAGAGACCGGCCCAAGGCCCAAC 300
Qy 301 AGCCGAGAGCTGAGAGTGCAGGCGCGCAACCCCGGAGAGGCGCGCGCGAGCGCGCAG 360
Db 301 AGCCGAGAGCTGAGAGTGCAGGCGCGCAACCCCGGAGAGGCGCGCGCGAGCGCGCAG 360
Qy 361 GGCACCTGAACTTCCCGCAGATCACTGTGAGCAGCGCCCTGTGAGCATCAAGT 420
Db 361 GGCACCTGAACTTCCCGCAGATCACTGTGAGCAGCGCCCTGTGAGCATCAAGT 420
Qy 421 GGGCGGCAAGTCAAGAGGCGCTGTGAGCAACGCGCGCGAGCAACCGTGTGAGAG 480
Db 421 GGGCGGCAAGTCAAGAGGCGCTGTGAGCAACGCGCGCGAGCAACCGTGTGAGAG 480
Qy 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGAGGCGGCACTGCGGCTTCAACAG 540
Db 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGAGGCGGCACTGCGGCTTCAACAG 540
Qy 541 GTGCGGCAAGTCAAGAGTCTGATCGAGATCTGCGGCAAGAGGCACTGCGGCGT 600
Db 541 GTGCGGCAAGTCAAGAGTCTGATCGAGATCTGCGGCAAGAGGCACTGCGGCGT 600
Qy 601 CTGATGCGGCGCCCAACCCCGTGAACATCATGCGCGCAACATGCTGAGCCGAGCTG 660
Db 601 CTGATGCGGCGCCCAACCCCGTGAACATCATGCGCGCAACATGCTGAGCCGAGCTG 660
Qy 661 ACCCTGAATCTCCCATCAAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGAG 720
Db 661 ACCCTGAATCTCCCATCAAGCCCATCGAGACCGTCCCGTGAAGCTGAAGCCCGGAG 720
Qy 721 GACGCGCCCAAGGATGAGAGGAGGCGCAAGATCAAGAGGCTGAGCGCC 780
Db 721 GACGCGCCCAAGGATGAGAGGAGGCGCAAGATCAAGAGGCTGAGCGCC 780
Qy 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAAGATCGGCCCGGAGAACCCCTAC 840
Db 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAAGATCGGCCCGGAGAACCCCTAC 840
Qy 841 AACACCCCGGTGTTCCCATCAAGAGAGAGAGCAACGAGTGGCGGCAAGCTGTGAG 900
Db 841 AACACCCCGGTGTTCCCATCAAGAGAGAGAGCAACGAGTGGCGGCAAGCTGTGAG 900
Qy 901 TTCCGAGAGCTGAACAGAGCAACCGAGACTTCTGAGAGGTGAGCTGAGCATCCCCAC 960
Db 901 TTCCGAGAGCTGAACAGAGCAACCGAGACTTCTGAGAGGTGAGCTGAGCATCCCCAC 960
Qy 961 CCGCGGCGCTGAAGAGAGAGAGAGGCTGAGCGTGTGAGCGTGGCGAGCGCTACTTC 1020
Db 961 CCGCGGCGCTGAAGAGAGAGAGAGGCTGAGCGTGTGAGCGTGGCGAGCGCTACTTC 1020
Qy 1021 AGGTGCGCTTGAACGAGAGCTTCCGAGATCAACGCGCTTCAACATCCCCAGATCAAC 1080
Db 1021 AGGTGCGCTTGAACGAGAGCTTCCGAGATCAACGCGCTTCAACATCCCCAGATCAAC 1080
Qy 1081 AACGAGACCCCGGAGATCGGCTACAGTACAGTGTGCTCCAGAGGCTGAGAGGAGC 1140
Db 1081 AACGAGACCCCGGAGATCGGCTACAGTACAGTGTGCTCCAGAGGCTGAGAGGAGC 1140
Qy 1141 CCGAGCATCTTCCAGAGAGAGATCAAGATCTGAGGCGCTTCCGCGCGCGCAACCC 1200
Db 1141 CCGAGCATCTTCCAGAGAGAGATCAAGATCTGAGGCGCTTCCGCGCGCGCAACCC 1200
Qy 1201 GAATGCTGATCTACAGATCAAGAGAGAGCTGATCTGAGGCGGAGCTGAGATGAGC 1260
Db 1201 GAATGCTGATCTACAGATCAAGAGAGAGCTGATCTGAGGCGGAGCTGAGATGAGC 1260

Db 1201 GAGATGATCTACCA-----GGCCCCtGTGTAAGTGGGAGAGCACTGAGATCGGC 1254
Qy 1261 CAGCACCGGCGCAAGATGAGAGAGTGTGCGCAAGCACTGTGCGCTGGGGCTTCAACACC 1320
Db 1255 CAGCACCGGCGCAAGATGAGAGAGTGTGCGCAAGCACTGTGCGCTGGGGCTTCAACACC 1314
Qy 1321 CCGGACAGAGAGCAAGAGAGAGGCGCGCTTCTGTGAGTGGGGCTACAGAGTGCACCC 1380
Db 1315 CCGGACAGAGAGCAAGAGAGAGGCGCGCTTCTGTGAGTGGGGCTACAGAGTGCACCC 1374
Qy 1381 GACAAATGAGACCGTCAAGCCCATCGAGTGTCCCGAGAGAGAGAGTGAACGAC 1440
Db 1375 GACAAATGAGACCGTCAAGCCCATCGAGTGTCCCGAGAGAGAGAGTGAACGAC 1434
Qy 1441 ATCCAGAGCTGTGGCAAGCTGAACCTGGGCAAGCAGATCTTACCCGCAATCAAGT 1500
Db 1435 ATCCAGAGCTGTGGCAAGCTGAACCTGGGCAAGCAGATCTTACCCGCAATCAAGT 1494
Qy 1501 CGCCAGCTGTGCAAGCTGTGCGCGCGCGCGCAAGGCGCTGACAGCACTGTCGCCAGCC 1560
Db 1495 CGCCAGCTGTGCAAGCTGTGCGCGCGCGCGCAAGGCGCTGACAGCACTGTCGCCAGCC 1554
Qy 1561 GAGAGGCGGAGCTGAGAGCTGCGCGGAGAACCGGAGATCTGCGGAGCGCGTGCACGCG 1620
Db 1555 GAGAGGCGGAGCTGAGAGCTGCGCGGAGAACCGGAGATCTGCGGAGCGCGTGCACGCG 1614
Qy 1621 GTGTACTAGACCCCGAGAGAGACTGTGGCCGAGATTCAGAGAGAGGCGCACAGCAG 1680
Db 1615 GTGTACTAGACCCCGAGAGAGACTGTGGCCGAGATTCAGAGAGAGGCGCACAGCAG 1674
Qy 1681 TGGACTTACAGATCTACAGAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAG 1740
Db 1675 TGGACTTACAGATCTACAGAGAGCGCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAG 1734
Qy 1741 ATGCGCACCGGCGCAACCAAGAGAGAGAGCTGAGAGAGAGAGAGAGATCGCC 1800
Db 1735 ATGCGCACCGGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGATCGCC 1794
Qy 1801 ATGAGAGAGATGATGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 1795 ATGAGAGAGATGATGATCTGTGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Qy 1861 ACTTGGAGAGAGCTGTGTGAGACCGAGCTACAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 1855 ACTTGGAGAGAGAGCTGTGTGAGACCGAGCTACAGAGAGAGAGAGAGAGAGAGAG 1914
Qy 1921 GTGAACACCCCGCGTGTGAGAGCTGTGTGAGAGCTGTGAGAGAGAGAGAGAGAGAG 1980
Db 1915 GTGAACACCCCGCGTGTGAGAGCTGTGTGAGAGCTGTGAGAGAGAGAGAGAGAGAG 1974
Qy 1981 GCGGAGACCTTCTACGTGAGAGCGGCGCGCAACCGGAGAGCAAGATCGGAGAGCGCGC 2040
Db 1975 GCGGAGACCTTCTACGTGAGAGCGGCGCGCAACCGGAGAGCAAGATCGGAGAGCGCGC 2034
Qy 2041 TACGTGAG 2100
Db 2035 TACGTGAG 2094
Qy 2101 ACCGAGCTGAG 2160
Db 2095 ACCGAGCTGAG 2154
Qy 2161 ACCGAG 2220
Db 2155 ACCGAG 2214
Qy 2221 CTGTGAG 2280
Db 2215 CTGTGAG 2274
Qy 2281 CCGGCGCAAG 2340
Db 2275 CCGGCGCAAG 2334

QY 2341 CGCAAGTCTCTTCTCTGAGCGGCATTCATGATGATCTACCAAGTACATGAGC 2400
| | | | |
DB 2335 CGCAAGTCTCTTCTCTGAGCGGCATTCATGATGATCTACCAAGTACATGAGC 2394
| | | | |
QY 2401 GACCTGTACGTGGGAGCGCGGCTTACGATGATTAAGCTTCCCGGGCTAGCAC 2460
| | | | |
DB 2395 GACCTGTACGTGGGAGCGCGGCTTACGATGATTAAGCTTCCCGGGCTAGCAC 2454
| | | | |
QY 2461 GGT 2463
| | | | |
DB 2455 GGT 2457
| | | | |
RESULT 9
ABL39961
ID ABL39961 standard; DNA; 2457 BP.
XX ABL39961;
AC
XX
XX ABL39961;
DT 15-MAY-2002 (first entry)
XX
XX
DE Synthetic construct PR975YMM SEQ ID NO:32.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; de.
XX
XX Human immunodeficiency virus; type C.
OS Synthetic.
OS
PN WO200204493-A2.
XX
XX 17-JAN-2002.
PD
XX 05-JUL-2001; 2001WO-US021241.
PF
XX 05-JUL-2000; 2000US-00610313.
PR
XX
XX (CHTR) CHIRON CORP.
PA (UYST-) UNIV STELLERBOSCH.
XX
XX Zur Megele J, Barnett SM, Engelbrecht S, Van Rensburg EJ;
PI
XX
XX WPI; 2002-154920/20.
DR
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunisation or generation of packaging cell
lines, particularly in gene therapy.
XX
XX
XX Claim 1; Fig 10; 233pp; English.
XX
XX The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (i) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;
Query Match 97.8%; Score 2415.4; DB 6; Length 2457;
Best Local Similarity 99.3%; Pred. No. 3.7e-291;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 61 CGCAGCACTTCAAGGCCCCAAGGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGC 120
| | | | |
DB 61 CGCAGCACTTCAAGGCCCCAAGGCATCATCAAGTCTTCAACTGCGGCAAGAGGCGC 120
| | | | |
QY 121 CACATGCGCCGCAACTGCGCGCCCCCGCAAGAGGCTGTGAAAGTGTGCGCAAGAG 180
| | | | |
DB 121 CACATGCGCCGCAACTGCGCGCCCCCGCAAGAGGCTGTGAAAGTGTGCGCAAGAG 180
| | | | |
QY 181 GGCACCAAGATGAAGAGACTGCAACCGAGCGCAAGCTTCTTCCGGAAGGACTTGGCC 240
| | | | |
DB 181 GGCACCAAGATGAAGAGACTGCAACCGAGCGCAAGCTTCTTCCGGAAGGACTTGGCC 240
| | | | |
QY 241 TTCCCCAAGGAGCCCGCGAGTTCCCCAGCGAGCAAGAACCGGCGCAACAGCCCCACC 300
| | | | |
DB 241 TTCCCCAAGGAGCCCGCGAGTTCCCCAGCGAGCAAGAACCGGCGCAACAGCCCCACC 300
| | | | |
QY 301 AGCGCGAGCTGACAGTGTGCGCGCAACCCCGGACGAGGCGCGGCGCGAGCGCGAG 360
| | | | |
DB 301 AGCGCGAGCTGACAGTGTGCGCGCAACCCCGGACGAGGCGCGGCGCGAGCGCGAG 360
| | | | |
QY 361 GGCACCTTGAATTTCCCCCAATCACTCTGTGAGCGGCCCCCTGTGTGATCAAGGTG 420
| | | | |
DB 361 GGCACCTTGAATTTCCCCCAATCACTCTGTGAGCGGCCCCCTGTGTGATCAAGGTG 420
| | | | |
QY 421 GCGCGCAGATCAAGAGAGCCCTGTGAGCAACCGGCGCGAGCAACCGGTGTGAGAG 480
| | | | |
DB 421 GCGCGCAGATCAAGAGAGCCCTGTGAGCAACCGGCGCGAGCAACCGGTGTGAGAG 480
| | | | |
QY 481 ATGAGCTGTCCCGGCAAGTGAAGCCCAAGATGATGAGCGGATGAGCGGCTTCATCAAG 540
| | | | |
DB 481 ATGAGCTGTCCCGGCAAGTGAAGCCCAAGATGATGAGCGGATGAGCGGCTTCATCAAG 540
| | | | |
QY 541 GTGCGCCAGTACAGACCAATCTGATTCGAGATCTGCGGCAAGAGCCCATGCGCACCGTG 600
| | | | |
DB 541 GTGCGCCAGTACAGACCAATCTGATTCGAGATCTGCGGCAAGAGCCCATGCGCACCGTG 600
| | | | |
QY 601 CTGATCGGCCCCACCCCGGTGACATCATCGGCGGCAACATGCTGACCCGAGCTGCGC 660
| | | | |
DB 601 CTGATCGGCCCCACCCCGGTGACATCATCGGCGGCAACATGCTGACCCGAGCTGCGC 660
| | | | |
QY 661 ACCCTGAATCTTCCCATGAGCCCATGAGACCGGTGCGGTGAGAGCTGAGACCGGCGATG 720
| | | | |
DB 661 ACCCTGAATCTTCCCATGAGCCCATGAGACCGGTGCGGTGAGAGCTGAGACCGGCGATG 720
| | | | |
QY 721 GACGCGCCCAAGGTGAAGAGTGTGAGCCCTGACCCGAGAGAGATCAAGGCTTGAACCGC 780
| | | | |
DB 721 GACGCGCCCAAGGTGAAGAGTGTGAGCCCTGACCCGAGAGAGATCAAGGCTTGAACCGC 780
| | | | |
QY 781 ATCTCGAGAGATGAGAGAGGAGCAAGATCAACCAAGATGAGGCCCCGAGAACCCCTTAC 840
| | | | |
DB 781 ATCTCGAGAGATGAGAGAGGAGCAAGATCAACCAAGATGAGGCCCCGAGAACCCCTTAC 840
| | | | |
QY 841 AACACCCCGTGTGTCGATCAAGAGAGAGCAGACCAAGTGTGCGGCAAGTGTGAGC 900
| | | | |
DB 841 AACACCCCGTGTGTCGATCAAGAGAGAGCAGACCAAGTGTGCGGCAAGTGTGAGC 900
| | | | |
QY 901 TTTCGCGAGCTGAACAAGCGCACCGAGACTTGTGAGAGTGTGACTGTGGCATTCGCCAC 960
| | | | |
DB 901 TTTCGCGAGCTGAACAAGCGCACCGAGACTTGTGAGAGTGTGACTGTGGCATTCGCCAC 960
| | | | |
QY 961 CCGCGCGGCTGAAGAAGAGAGAGCGTGAACCGGTGAGAGCTGTGAGAGCGCTTACTTTC 1020
| | | | |
DB 961 CCGCGCGGCTGAAGAAGAGAGAGCGTGAACCGGTGAGAGCTGTGAGAGCGCTTACTTTC 1020
| | | | |
QY 1021 AGCGTGCCTGTGAAGAGACTTTCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAAC 1080
| | | | |
DB 1021 AGCGTGCCTGTGAAGAGACTTTCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAAC 1080
| | | | |
QY 1081 AACGAGACCCCGGCGATCGCTTCAAGTACAACTGTGCGCCGAGGCTGGAAGGCGAGC 1140
| | | | |
DB 1081 AACGAGACCCCGGCGATCGCTTCAAGTACAACTGTGCGCCGAGGCTGGAAGGCGAGC 1140
| | | | |

QY 1141 CCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGAGACCTTCCGGGCGGCAACCC 1200
 DB 1141 CCCAGCATCTTTCAGAGCAGCATGACCAAGATCTTGAGACCTTCCGGGCGGCAACCC 1200
 QY 1201 GAGATCGATCTTACAGATGACATGACAGCCTGTGCTGTGGGAGGAGCCTTGAGATCGGC 1260
 DB 1201 GAGATCGATCTTACCA-----GGCCCCCTGTGCTGTGGGAGGAGCCTTGAGATCGGC 1254
 QY 1261 CAGCACCGGCGCAGATCGAGAGGCTGCGCAAGCCTTGCTGCGCTGGGGCTTCAACAC 1320
 DB 1255 CAGCACCGGCGCAGATCGAGAGGCTGCGCAAGCCTTGCTGCGCTGGGGCTTCAACAC 1314
 QY 1321 CCCGACAAAGACACCAAGAGAGCCCCCTTCTGTGATGAGGCTACGAGCTTGACCCC 1380
 DB 1315 CCCGACAAAGACACCAAGAGAGCCCCCTTCTGTGATGAGGCTACGAGCTTGACCCC 1368
 QY 1381 GACCAAGTGGACCGTGAAGCCCATTCGAGCTGCGGAGAGAGAGTGGACCGTGAACGAC 1440
 DB 1369 GACCAAGTGGACCGTGAAGCCCATTCGAGCTGCGGAGAGAGAGTGGACCGTGAACGAC 1428
 QY 1441 ATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGAGCCAGATCTACCCCGGCATCAAGTG 1500
 DB 1429 ATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGAGCAGATCTACCCCGGCATCAAGTG 1488
 QY 1501 CCGCAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCCCTTGAACGACATCTGTCCCTGACC 1560
 DB 1489 CCGCAGCTGTGCAAGCTGCTGCGCGGCGCAAGGCCCTTGAACGACATCTGTCCCTGACC 1548
 QY 1561 GAGGAGCGGCAAGCTGAGAGTGGCCGAGAACCGCGAGATCTGCGGAGGCCGCTGACCGGC 1620
 DB 1549 GAGGAGCGGCAAGCTGAGAGTGGCCGAGAACCGCGAGATCTGCGGAGGCCGCTGACCGGC 1608
 QY 1621 GTGTACTAGCAACCCAGCAAGGACCTGTGCGCGAGATCCAGAAAGAGGCGCAACGACG 1680
 DB 1609 GTGTACTAGCAACCCAGCAAGGACCTGTGCGCGAGATCCAGAAAGAGGCGCAACGACG 1668
 QY 1681 TGGACCTTACAGATCTTACAGAGACCTTTCAGAAACCTTGAAGACCGGCAAGTACGCGCAG 1740
 DB 1669 TGGACCTTACAGATCTTACAGAGACCTTTCAGAAACCTTGAAGACCGGCAAGTACGCGCAG 1728
 QY 1741 ATGCGCAGCGGCCCAACCAAGCAGTGAAGAGCTGACGAGAGGCCGTGAGAGATCGGC 1800
 DB 1729 ATGCGCAGCGGCCCAACCAAGCAGTGAAGAGCTGACGAGAGGCCGTGAGAGATCGGC 1788
 QY 1801 ATGAGAGCATCGTGTCTGTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGAGAG 1860
 DB 1789 ATGAGAGCATCGTGTCTGTGGGCAAGACCCCAAGTTCGCGCTGCCCATCCAGAGAGAG 1848
 QY 1861 ACCCTGAGAGACCTGTGAGACCGACTGCTGAGAGGCCACCTGATCCCGAGTGGAGTTTC 1920
 DB 1849 ACCCTGAGAGACCTGTGAGACCGACTGCTGAGAGGCCACCTGATCCCGAGTGGAGTTTC 1908
 QY 1921 GTGAAACACCCCCCTGTGTGAAGCTGTGTGATCCACTGTGAGAGAGAGCCCATCATCGGC 1980
 DB 1909 GTGAAACACCCCCCTGTGTGAAGCTGTGTGATCCACTGTGAGAGAGAGCCCATCATCGGC 1968
 QY 1981 GCCGAGACCTTCTAGTGTGAGCGGCGCGCAACCGGAGAGCAAGATCCGCAAGGCGGCG 2040
 DB 1969 GCCGAGACCTTCTAGTGTGAGCGGCGCGCAACCGGAGAGCAAGATCCGCAAGGCGGCG 2028
 QY 2041 TACGTGACCGACCGGGGCGGCGAGAAAGATGTGAGCCCTGACCGAGACCAACCAAGAG 2100
 DB 2029 TACGTGACCGACCGGGGCGGCGAGAAAGATGTGAGCCCTGACCGAGACCAACCAAGAG 2088
 QY 2101 ACCGAGCTGAGAGCCATTCACGCTGGCCCTGACAGACAGCGGCGCGAGAGTGAACATCTG 2160
 DB 2089 ACCGAGCTGAGAGCCATTCACGCTGGCCCTGACAGACAGCGGCGCGAGAGTGAACATCTG 2148
 QY 2161 ACCGAGCGCAGTACGCGCTGGGCGATCATCAGAGCCAGCGGCAAGAGCGAGAGCGAG 2220
 DB 2149 ACCGAGCGCAGTACGCGCTGGGCGATCATCAGAGCCAGCGGCAAGAGCGAGAGCGAG 2208
 QY 2221 CTGTGAAACCAAGATCATCGAGAGCTGATCAAGAGAGAGGTGTACTTGAGCTGGGTG 2280

DB 2209 CTGTGAAACCAAGATCATCGAGAGCTGTATCAAGAGAGAGGTGTACTGAGCTGGGTG 2268
 QY 2281 CCCGCCCAAGAGGATCGGCGGCAACGAGATCGAACAGCTGTGAGAGAGGCGATC 2340
 DB 2269 CCCGCCCAAGAGGATCGGCGGCGCAACGAGATCGAACAGCTGTGAGAGGCGATC 2328
 QY 2341 CGCAAGGTGCTGTCTTGTGAGCGGCATCGATGGCGGCATCGATCTTACCAAGTGAAC 2400
 DB 2329 CGCAAGGTGCTGTCTTGTGAGCGGCATCGATGGCGGCATCGATCTTACCAAGTGAAC 2388
 QY 2401 GACCTGTAGTGGGAGGCGCGCCCTAGATCGATTAAAGCTTCCGCGGCTGACACC 2460
 DB 2389 GACCTGTAGTGGGAGGCGCGCCCTAGATCGATTAAAGCTTCCGCGGCTGACACC 2448
 QY 2461 GGTGAATTC 2469
 DB 2449 GGTGAATTC 2457
 RESULT 10
 ADM73766
 ID ADM73766 strand; DNA; 2457 BP.
 XX
 AC ADM73766;
 XX
 DT 03-JUN-2004 (first entry)
 XX
 DE HIV-1 polynucleotide #9.
 XX
 KW HIV-1; gene; ds; HIV pol; immune response; DNA immunisation;
 KW HIV type C protein; immunostimulant.
 OS Human immunodeficiency virus 1.
 XX
 PN US2003223961-A1.
 XX
 PD 04-DEC-2003.
 XX
 PF 05-JUL-2001; 2001US-00899575.
 PR 05-JUL-2000; 2000US-00610313.
 XX
 PA (MEGEDE J Z.
 PA (BARV/) BARNETT S W.
 PA (ENG/) ENGELBRECHT S.
 PA (RENS/) RENSBURG B J V.
 XX
 PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;
 XX
 DR WPI; 2004-060515/06.
 XX
 PT New expression cassette comprising a polynucleotide sequence encoding an
 PT HIV Pol polypeptide, useful in eliciting an immune response, in DNA
 PT immunization, generating of packaging cell lines or in producing HIV type
 PT C proteins.
 PS Claim 1; SEQ ID NO 32; 160pp; English.
 XX
 CC The invention relates to an expression cassette comprising a
 CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention
 CC also relates to a recombinant expression system for use in a host cell
 CC comprising an expression cassette, where the polynucleotide sequence
 CC further comprises control elements capable of driving expression in the
 CC selected host cell, a cell comprising an expression cassette where the
 CC polynucleotide sequence further comprises control elements compatible
 CC with the expression in the cell and a composition for generating an
 CC immunological response, comprising an expression cassette. The expression
 CC cassette and the methods of the invention are useful in eliciting an
 CC immune response, in DNA immunisation, in generation of packaging cell
 CC lines and in producing HIV type C proteins. This sequence represents an
 CC HIV-1 polynucleotide of the invention.

Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 97.8%; Score 2415.4; DB 12; Length 2457;
 Best Local Similarity 99.3%; Pred. No. 3.7e-291;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 1 GTGACGCGCACCATGAGCGCGACATGAGCCAGGCGCACCGAGCCCAACATCTCTGATGCG 60
 DB 1 GTGACGCGCACCATGAGCGCGACATGAGCCAGGCGCACCGAGCCCAACATCTCTGATGCG 60
 QY 61 CGCAGCACTTTCAAGAGGCGCGCAAGGCGCATATCAAGTGTCTTCAACTGCGGCGCAAGAGGCG 120
 DB 61 CGCAGCACTTTCAAGAGGCGCGCAAGGCGCATATCAAGTGTCTTCAACTGCGGCGCAAGAGGCG 120
 QY 121 CACATGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGGCGGCAAGAG 180
 DB 121 CACATGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGGCGGCAAGAG 180
 QY 181 GCGCAGCAGATGAGAGCTGCAACCGAGCGCGCAAGCCTTCTTCCGCGAGGACCTTGGCC 240
 DB 181 GCGCAGCAGATGAGAGCTGCAACCGAGCGCGCAAGCCTTCTTCCGCGAGGACCTTGGCC 240
 QY 241 TTCCCGCAGGCGCAAGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCGCAACGCGCCACC 300
 DB 241 TTCCCGCAGGCGCAAGGCGCGCGAGTTCCCGAGCGAGCAGAACCGCGCGCAACGCGCCACC 300
 QY 301 AGCGCGAGCTGCAAGTGGCGCGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 360
 DB 301 AGCGCGAGCTGCAAGTGGCGCGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCG 360
 QY 361 GCGACCTTGAACTTCCCGCGAGTCAACCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 420
 DB 361 GCGACCTTGAACTTCCCGCGAGTCAACCTGTGAGAGCGCGCGCGCGCGCGCGCGCGCGCG 420
 QY 421 GCGCGCGCAGATTAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 DB 421 GCGCGCGCAGATTAAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
 QY 481 ATGAGCTGCG 540
 DB 481 ATGAGCTGCG 540
 QY 541 GTGCGCGAGTACAGCAGATCTGTATCGAGATCTGCGCGAGAGAGCGCATTCGCGACCGTG 600
 DB 541 GTGCGCGAGTACAGCAGATCTGTATCGAGATCTGCGCGAGAGAGCGCATTCGCGACCGTG 600
 QY 601 CTGATTCG 660
 DB 601 CTGATTCG 660
 QY 661 ACCCTGAACCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
 DB 661 ACCCTGAACCTTCCCGCATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
 QY 721 GACG 780
 DB 721 GACG 780
 QY 781 ATCTCGAGAGAGATGAGAGAGAGAGAGAGAGATCAACAGATCGCGCGCGCGCGCGCGCG 840
 DB 781 ATCTCGAGAGAGATGAGAGAGAGAGAGAGAGATCAACAGATCGCGCGCGCGCGCGCGCG 840
 QY 841 AACACCG 900
 DB 841 AACACCG 900
 QY 901 TTCCCGAGCTGAAACAAGCGCACCGAGACTTCTGAGAGGTGCAAGTGGCGCAACCGCCAC 960
 DB 901 TTCCCGAGCTGAAACAAGCGCACCGAGACTTCTGAGAGGTGCAAGTGGCGCAACCGCCAC 960
 QY 961 CCGCGCGCGCTGAAAG 1020
 DB 961 CCGCGCGCGCTGAAAG 1020

QY 1021 AGCGTGGCGCGTGAAG 1080
 DB 1021 AGCGTGGCGCGTGAAG 1080
 QY 1081 AACGAGAGCG 1140
 DB 1081 AACGAGAGCG 1140
 QY 1141 CCGAGCATCTTCCAG 1200
 DB 1141 CCGAGCATCTTCCAG 1200
 QY 1201 GAGATCGAGATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 DB 1201 GAGATCGAGATCTACAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1261 CAGCAGCG 1320
 DB 1261 CAGCAGCG 1320
 QY 1321 CCGGAG 1380
 DB 1321 CCGGAG 1380
 QY 1381 GACAG 1440
 DB 1381 GACAG 1440
 QY 1441 ATCCAG 1500
 DB 1441 ATCCAG 1500
 QY 1501 GCGCAG 1560
 DB 1501 GCGCAG 1560
 QY 1561 GAGGAG 1620
 DB 1561 GAGGAG 1620
 QY 1621 GTGTACTACAG 1680
 DB 1621 GTGTACTACAG 1680
 QY 1681 TGGACCTACAGAGATTTACAG 1740
 DB 1681 TGGACCTACAGAGATTTACAG 1740
 QY 1741 ATGCGCAGCG 1800
 DB 1741 ATGCGCAGCG 1800
 QY 1801 ATGAGAGCATGTGATCTGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 DB 1801 ATGAGAGCATGTGATCTGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 1861 ACTTGGAG 1920
 DB 1861 ACTTGGAG 1920
 QY 1921 GTGAAACAGCG 1980
 DB 1921 GTGAAACAGCG 1980
 QY 1981 GCGGAG 2040
 DB 1981 GCGGAG 2040
 QY 2041 TACGTGAG 2100
 DB 2041 TACGTGAG 2100
 QY 2088 TACGTGAG 2088
 DB 2088 TACGTGAG 2088

QY 2101 ACCGAGCTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2160
DB 2089 ACCGAGCTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2148
QY 2161 ACCGAGCTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2220
DB 2149 ACCGAGCTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2208
QY 2221 CTGTGTGAACCAATCATCATGACAGCTGATCATAGAGAGAGAGGTGTACTGAGCTGGTG 2280
DB 2209 CTGTGTGAACCAATCATCATGACAGCTGATCATAGAGAGAGAGGTGTACTGAGCTGGTG 2268
QY 2281 CCGGCGCAAGAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2340
DB 2269 CCGGCGCAAGAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2328
QY 2341 CCGAAGGTCTGTCTTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2400
DB 2329 CCGAAGGTCTGTCTTGACAGGCAATCCAGCTGCGCTTGACAGACAGGCGGAGGTTGAACATCTG 2388
QY 2401 GACCTGTACGTGGGAGGCGGCGGCTTGAAGATCATTAAGCTTCCGCGGCTAGCAC 2460
DB 2389 GACCTGTACGTGGGAGGCGGCGGCTTGAAGATCATTAAGCTTCCGCGGCTAGCAC 2448
QY 2461 GGTGAATTC 2469
DB 2449 GGTGAATTC 2457

RESULT 11
ID ACA03546 standard; DNA; 2445 BP.
AC ACA03546;
AC 22-MAY-2003: (first entry)
DT Synthetic DNA encoding immunogenic HIV peptide #29.
DE Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; ds.
XX
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megele J, Barnett SW, Lian Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 34; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for

CC immunisation, generation of packaging cell line, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;
QY Query Match 97.3%; Score 2401.8; DB 8; Length 2445;
DB Best Local Similarity 99.2%; Pred. No. 1.8e-289;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;
QY 7 GCCACCATGGCCGAGGCAATGAGCCAGGCGGCAAGGCTTGTGACAGCCAGC 66
DB 1 GCCACCATGGCCGAGGCAATGAGCCAGGCGGCAAGGCTTGTGACAGCCAGC 60
QY 67 AACTTCAGAGGCGCCCAAGGCAATCATCATGCTTCACTGCGGCAAGAGGCGCATC 126
DB 61 AACTTCAGAGGCGCCCAAGGCAATCATCATGCTTCACTGCGGCAAGAGGCGCATC 120
QY 127 GCCCGCACTGCGCGGCGCCCGGCAAGAGGCGTGTGAGAGTGGCGCAAGAGGCGCAC 186
DB 121 GCCCGCACTGCGCGGCGCCCGGCAAGAGGCGTGTGAGAGTGGCGCAAGAGGCGCAC 180
QY 187 CAGATGAAGATGTCACACGAGGCGGCAAGGCTTTCGCGAGGACCTTGCTTCC 246
DB 181 CAGATGAAGATGTCACACGAGGCGGCAAGGCTTTCGCGAGGACCTTGCTTCC 240
QY 247 CAGGCGAAGCGCGCGAGTTCCTCCAGGCAAGAGGCGGCAAGGCGGCGGCGGCGG 306
DB 241 CAGGCGAAGCGCGCGAGTTCCTCCAGGCAAGAGGCGGCGGCGGCGGCGGCGGCGG 300
QY 307 GAGCTGAGAGTGGCGGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 366
DB 301 GAGCTGAGAGTGGCGGCGGCAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 367 CTGAATCTTCCCGCAGATCAACCTGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 426
DB 361 CTGAATCTTCCCGCAGATCAACCTGTGTGAGGCGGCGGCGGCGGCGGCGGCGGCGG 420
QY 427 CAGATGAAGAGGCGGCTGTGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 486
DB 421 CAGATGAAGAGGCGGCTGTGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 487 CTGCGCGGCAAGTGAAGGCGGCAAGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGG 546
DB 481 CTGCGCGGCAAGTGAAGGCGGCAAGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 547 CAGTACGACAGATCTGTATCGAGATCTGCGGCAAGAGGCGGCGGCGGCGGCGGCGG 606
DB 541 CAGTACGACAGATCTGTATCGAGATCTGCGGCAAGAGGCGGCGGCGGCGGCGGCGG 600
QY 607 GGGCCCAAGCGGCGGCGGCAATCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 666
DB 601 GGGCCCAAGCGGCGGCGGCAATCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 667 AACTTCCCATGAGGCGGCGGCAATCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 726
DB 661 AACTTCCCATGAGGCGGCGGCAATCATCATGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720
QY 727 CCGAAGTGAAGAGTGGCGGCGGCGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 786
DB 721 CCGAAGTGAAGAGTGGCGGCGGCGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 787 GAGAGATGAGAGAGGCGGCAAGATCATCATGAGTGGCGGCGGCGGCGGCGGCGGCGG 846
DB 781 GAGAGATGAGAGAGGCGGCAAGATCATCATGAGTGGCGGCGGCGGCGGCGGCGGCGG 840
QY 847 CCGGTGTGCGCATCAAGAGAGGCAAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 906

Db 841 CCCGTTCCCATCAAGAAAGAGACCAAGTGGCGAAGCTGTGACCTTCCGC 900
Qy 907 GAGCTGAACAAAGCCACCCAGAGCTTCTGAGAGTGCAGCTGGGCAATCCCCACCCCGC 966
Db 901 GAGCTGAACAAAGCCACCCAGAGCTTCTGAGAGTGCAGCTGGGCAATCCCCACCCCGC 960
Qy 967 GGCCTGAAGAAGAAAGAGCGTGAACCGTGTGAACGTGGGCGAAGCCTTCAAGCGTG 1026
Db 961 GGCCTGAAGAAGAAAGAGCGTGAACCGTGTGAACGTGGGCGAAGCCTTCAAGCGTG 1020
Qy 1027 CCCCTGAAGAGAGCTTCCGCAAGTACACCGCTTTCACCATCCCAAGATCAACAGAG 1086
Db 1021 CCCCTGAAGAGAGCTTCCGCAAGTACACCGCTTTCACCATCCCAAGATCAACAGAG 1080
Qy 1087 ACCCCCGGATCCGCTTACCAAGTACCAAGCTGTGCCCCAGAGGCTGAAGAGGCGCCAGC 1146
Db 1081 ACCCCCGGATCCGCTTACCAAGTACCAAGCTGTGCCCCAGAGGCTGAAGAGGCGCCAGC 1140
Qy 1147 ATCTTCCAGAGAGCATGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGAGCATGACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTACAGTACATGACGACCTGTACGTGGGAGCGACCTGAGATCCGCGAGAC 1266
Db 1201 GTGATCTACCA-----GGCCCCCTGTACGTGGGAGCGACCTGAGATCCGCGAGAC 1254
Qy 1267 CCGCGCAGATGAGAGAGCTGGCGAAGACCTGTGCGCTGGGCGTTCAACACCCCGAC 1326
Db 1255 CCGCGCAGATGAGAGAGCTGGCGAAGACCTGTGCGCTGGGCGTTCAACACCCCGAC 1314
Qy 1327 AAGAGACACAGAAAGAGCCCGCTTCTGTGAGTGGGCTAGAGCTGACCCCGACAG 1386
Db 1315 AAGAGACACAGAAAGAGCCCGCTTCTGTGAGTGGGCTAGAGCTGACCCCGACAG 1368
Qy 1387 TGGACCTGTGACGCCCATCGAGCTGCGCGAAGAGAGAGCTGACCGTGAACGACATCCAG 1446
Db 1369 TGGACCTGTGACGCCCATCGAGCTGCGCGAAGAGAGAGCTGACCGTGAACGACATCCAG 1428
Qy 1447 AAGCTGTGGGCAAGCTGAACTGGGCGACCGCAGATTTACCCCGCATCAAGTGGCGCAG 1506
Db 1429 AAGCTGTGGGCAAGCTGAACTGGGCGACCGCAGATTTACCCCGCATCAAGTGGCGCAG 1488
Qy 1507 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATGTGCCCCGACCGAGAG 1566
Db 1489 CTGTGCAAGCTGTGCGCGCGCGCGCAAGGCGCTGACCGACATGTGCCCCGACCGAGAG 1548
Qy 1567 GCCGAGCTGAGAGCTGGCCGAGAACCGCGAGATCTGTGCGGAGCCCGTGCACGCGCTGAC 1626
Db 1549 GCCGAGCTGAGAGCTGGCCGAGAACCGCGAGATCTGTGCGGAGCCCGTGCACGCGCTGAC 1608
Qy 1627 TACGACCCCGCAAGAGACTGTGTGGCGGAGATTCAGAAAGAGGCGACGACGATGAGACC 1686
Db 1609 TACGACCCCGCAAGAGACTGTGTGGCGGAGATTCAGAAAGAGGCGACGACGATGAGACC 1668
Qy 1687 TACGAGATCTACAGAGAGCCCTTCAAGACTGAAGACCGGCAAGTGAAGCGCAAGTGGC 1746
Db 1669 TACGAGATCTACAGAGAGCCCTTCAAGACTGAAGACCGGCAAGTGAAGCGCAAGTGGC 1728
Qy 1747 ACCGCGCACACCAAGAGAGCTGAGACGAGCGGTGACAGAGAGTGCATGAGAG 1806
Db 1729 ACCGCGCACACCAAGAGAGCTGAGACGAGCGGTGACAGAGAGTGCATGAGAG 1788
Qy 1807 AGCATGTGTATCTGGGGCAAGACCCCGAAGTTCCGCTGCGCATTCAGAAAGAGACTGG 1866
Db 1789 AGCATGTGTATCTGGGGCAAGACCCCGAAGTTCCGCTGCGCATTCAGAAAGAGACTGG 1848
Qy 1867 GAGACTGTGTGACCGACTACTGAGAGGCGACCTGGATCCCGAGTGGAGTTCGTGAAC 1926
Db 1849 GAGACTGTGTGACCGACTACTGAGAGGCGACCTGGATCCCGAGTGGAGTTCGTGAAC 1908
Qy 1927 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1986

Db 1909 ACCCCCCCTGTGTGAAGCTGTGTACAGCTGAGAGAGAGCCCATCATCGGCGCGAG 1968
Qy 1987 ACCTTCTACGTGAGACGGGCGCGCCCAACCGCGAAGACCAAGATGAGCAAGGCCCGGCTACGTG 2046
Db 1969 ACCTTCTACGTGAGACGGGCGCGCCCAACCGCGAAGACCAAGATGAGCAAGGCCCGGCTACGTG 2028
Qy 2047 ACCGACCGGGCGCGCAAGAGATCGTGAAGCTGACCGGAGACCAACCAAGAGAGCCGAG 2106
Db 2029 ACCGACCGGGCGCGCAAGAGATCGTGAAGCTGACCGGAGACCAACCAAGAGAGCCGAG 2088
Qy 2107 CTGCAAGCCATTCAGCTGGCCCTGTGAGAGACAGCGGCGAGGAGTGAACAATGTGACCAAC 2166
Db 2089 CTGCAAGCCATTCAGCTGGCCCTGTGAGAGACAGCGGCGAGGAGTGAACAATGTGACCAAC 2148
Qy 2167 AGCGATAGCGCTGTGGGATCATTCAGAGCCGACCGCGCAAGAGAGAGCGAGCGTGGT 2226
Db 2149 AGCGATAGCGCTGTGGGATCATTCAGAGCCGACCGCGCAAGAGAGAGCGAGCGTGGT 2208
Qy 2227 AACGATCATTCAGACAGCTGATCAAGAGAGAGTGTACTGAGCTGGGTGCGCGC 2286
Db 2209 AACGATCATTCAGACAGCTGATCAAGAGAGAGTGTACTGAGCTGGGTGCGCGC 2268
Qy 2287 CACAAAGGCAATCGCGGCAACGAGAGATCGAACAGCTGTGAGCAAGGCAATCCGCAAG 2346
Db 2269 CACAAAGGCAATCGCGGCAACGAGAGATCGAACAGCTGTGAGCAAGGCAATCCGCAAG 2328
Qy 2347 GTGCTGTTCTTGAAGACGACATGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2406
Db 2339 GTGCTGTTCTTGAAGACGACATGATGGCGGATCGTGTATCTACAGTACATGACGACCTG 2388
Qy 2407 TACGTGGGCAAGCGGCGCGCTGAGATCGATTTAAAGCTTCCCGGGCTAGACCGGT 2463
Db 2389 TACGTGGGCAAGCGGCGCGCTGAGATCGATTTAAAGCTTCCCGGGCTAGACCGGT 2445

RESULT 12
ADCl3264
ID ADCl3264 standard; DNA; 2445 BP.
XX
AC ADCl3264;
XX
DT 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct p2Pol-opt-YMM_C SEQ ID NO 43.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15Kaseh; Pol; Tat; Proct;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; da.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Llan Y, Engelbrecht S, Van Rensburg BJ,
XX
DR WPI; 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15Kaseh, Pol, Tat,
PT Proct, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 40; 301pp; English.

XX The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15maeh, Pol, Tat, Pro, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intramusosally, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 97.3%; Score 2401.8; DB 10; Length 2445;
Best Local Similarity 99.2%; Pred. No. 1.8e-289;
Matches 2438; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

QY 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCGCACCGCCCAACATCTGATGCAAGCGCAC 66
DB 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCGCACCGCCCAACATCTGATGCAAGCGCAC 60
QY 67 AACTTCAGAGGGGCCCAAGGCGCATCATCAAGTGTCTTAATGCGGCGAAGAGGGCCACATC 126
DB 61 AACTTCAGAGGGGCCCAAGGCGCATCATCAAGTGTCTTAATGCGGCGAAGAGGGCCACATC 120
QY 127 GCCCGCAATGCGCGCGCCCCCGCCAGAGAGGGCTGTGAGAGTGCAGGCGGCGGCGAC 186
DB 121 GCCCGCAATGCGCGCGCCCCCGCCAGAGAGGGCTGTGAGAGTGCAGGCGGCGGCGAC 180
QY 187 CAGATGAGAGATGTCACCGAGCGCCAGGCGCAATTTCTTCGCGAGGAGACTGTGCTTCC 246
DB 181 CAGATGAGAGATGTCACCGAGCGCCAGGCGCAATTTCTTCGCGAGGAGACTGTGCTTCC 240
QY 247 CAGGGCGAAGGCCCGCGAGTTTCCCGAGAGCGAGAACCGCGCCCAACGCCCCACACCC 306
DB 241 CAGGGCGAAGGCCCGCGAGTTTCCCGAGAGCGAGAACCGCGCCCAACGCCCCACACCC 300
QY 307 GAGCTGCAGATGCGCGGCGCAACCCCGCGAGGAGCGCGCGCGCGAGGCGCGAGGCGAC 366
DB 301 GAGCTGCAGATGCGCGGCGCAACCCCGCGAGGAGCGCGCGCGCGCGAGGCGCGAGGCGAC 360
QY 367 CTGAACTTCCCGCAGATCACTCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGGC 426
DB 361 CTGAACTTCCCGCAGATCACTCTGTGCGAGCGCCCTGTGTGAGCATCAAGGTGGCGGC 420
QY 427 CAGATCAAGAGAGGCGCTGTGTCACCGCGCGCGAGGAGCACTGTGTGAGAGATGAGC 486
DB 421 CAGATCAAGAGAGGCGCTGTGTCACCGCGCGCGAGGAGCACTGTGTGAGAGATGAGC 480
QY 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGAGTGGCGGCTTTCATCAAGGTGGC 546
DB 481 CTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGAGTGGCGGCTTTCATCAAGGTGGC 540
QY 547 CAGTACGACCAAGATCTGTATCGAGATCTGTGCGCAAGAGGCGCATCGGCACTGTATC 606
DB 541 CAGTACGACCAAGATCTGTATCGAGATCTGTGCGCAAGAGGCGCATCGGCACTGTATC 600
QY 607 GGGCCCAACCCCGGTGAACATCATCGGCGCGCAATGTGAACCAAGCTGGGCTCACCTTG 666
DB 601 GGGCCCAACCCCGGTGAACATCATCGGCGCGCAATGTGAACCAAGCTGGGCTCACCTTG 660
QY 667 AACTTCCCATCGAGCCCATCGAGCCGTGCGGTGAGTGAAGCCCGCATGGAAGCG 726
DB 661 AACTTCCCATCGAGCCCATCGAGCCGTGCGGTGAGTGAAGCCCGCATGGAAGCG 720
QY 727 CCCAAGGTGAAGAGTGGCCCTGACCGAGAGAGAAATCAAGGCCCTTGAACCGCATCTGC 786
DB 721 CCCAAGGTGAAGAGTGGCCCTGACCGAGAGAGAAATCAAGGCCCTTGAACCGCATCTGC 780
QY 787 GAGAGATGGAAGAGAGGCGCAAGATCAACAAATCGGCCCGAGAACTCTTCAACAC 846

DB 781 GAGAGATGGAAGAGAGGCGCAAGATCAACAAATCGGCCCGAGAACTCTTCAACAC 840
QY 847 CCCGTGTTGCCCATCAAGAGAGAGCAGCACCAAGTGGCGCAAGCTGTGATCTTCCGC 906
DB 841 CCCGTGTTGCCCATCAAGAGAGAGCAGCACCAAGTGGCGCAAGCTGTGATCTTCCGC 900
QY 907 GAGCTGAACAAGCGGACCCAGAGACTTCTGAGAGTGCAGCTGGGATATCCCGACCCGCG 966
DB 901 GAGCTGAACAAGCGGACCCAGAGACTTCTGAGAGTGCAGCTGGGATATCCCGACCCGCG 960
QY 967 GGCCTGAAGAGAGAGAGCCTGACCGGTGTGAGAGTGGGCGAGCGCTTACTTCAAGCTTG 1026
DB 961 GGCCTGAAGAGAGAGAGCCTGACCGGTGTGAGAGTGGGCGAGCGCTTACTTCAAGCTTG 1020
QY 1027 CCCCTGAGCAGAGACTTCCGCAAGTACACCGCTTTCACATCTCCAGCATCAACAGAG 1086
DB 1021 CCCCTGAGCAGAGACTTCCGCAAGTACACCGCTTTCACATCTCCAGCATCAACAGAG 1080
QY 1087 ACCCCCGGATCCGCTACCAAGTACACCGTGTGCCCGCGGCGTGAAGGGCGAGCCCGAC 1146
DB 1081 ACCCCCGGATCCGCTACCAAGTACACCGTGTGCCCGCGGCGTGAAGGGCGAGCCCGAC 1140
QY 1147 ATCTTCAGAGCAGATGACCAAGATCTGTAGAGCTTTCGCGCGCGCAACCCCGAGATC 1206
DB 1141 ATCTTCAGAGCAGATGACCAAGATCTGTAGAGCTTTCGCGCGCGCAACCCCGAGATC 1200
QY 1207 GTGATCTACAGATGATGAGACGACTGTACGTTGAGAGCGGACCTGAGATCGGCGACAC 1266
DB 1201 GTGATCTACG-----GGCCCCCTGTACGTGGGCGAGACTTGAATCGGCGAGCAC 1254
QY 1267 CCGCGCAAGATGAGAGAGTGTGCGCAAGACCTGTCTGCGTGGGGCTTCAACACCCCGAC 1326
DB 1255 CCGCGCAAGATGAGAGAGTGTGCGCAAGACCTGTCTGCGTGGGGCTTCAACACCCCGAC 1314
QY 1327 AAGAGCACAAGAGAGGCCCCCTTCTGTGATAGGGCTTACGAGCTGACCCCGACAAG 1386
DB 1315 AAGAGCACAAGAGAGGCCCCCTTCTGTGATAGGGCTTCTGTGATAGGGCTTCTGTGATAG 1368
QY 1387 TGAACCTGTGACGACCTTCAGAGCTGCGGAGAGAGAGAGTGTGACCTGTGAACGATCGAG 1446
DB 1369 TGAACCTGTGACGACCTTCAGAGCTGCGGAGAGAGAGAGTGTGACCTGTGAACGATCGAG 1428
QY 1447 AAGCTGTGTGGCAAGCTTGAACCTGTGTGCGGACCGAGATTTACCCCGGATCAAGGTGGCG 1506
DB 1429 AAGCTGTGTGGCAAGCTTGAACCTGTGTGCGGACCGAGATTTACCCCGGATCAAGGTGGCG 1488
QY 1507 CTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGACCGGACATGCTGTGCGGCGGAG 1566
DB 1489 CTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGACCGGACATGCTGTGCGGCGGAG 1548
QY 1567 GCCGAGCTGAGAGCTGTGCGGAGACCGCGAGATCTGTGCGGAGCGCGGTGTAC 1626
DB 1549 GCCGAGCTGAGAGCTGTGCGGAGACCGCGAGATCTGTGCGGAGCGCGGTGTAC 1608
QY 1627 TACGACCCCAAGAGAGCTGTGTGCGGAGATCTGAGAGCAAGGCGCAAGCTGTGAGCC 1686
DB 1609 TACGACCCCAAGAGAGCTGTGTGCGGAGATCTGAGAGCAAGGCGCAAGCTGTGAGCC 1668
QY 1687 TACGAGTCTTACAGAGAGGCTTTCAGAGACTTGAAGACCGGCAAGTACGCAAGATGGC 1746
DB 1669 TACGAGTCTTACAGAGAGGCTTTCAGAGACTTGAAGACCGGCAAGTACGCAAGATGGC 1728
QY 1747 ACCGCGCAACCAAGAGAGCTGTGAGAGCTGTGACCGGAGCGGTGCAAGATTCGCATGTAG 1806
DB 1729 ACCGCGCAACCAAGAGAGCTGTGAGAGAGCTGTGACCGGAGCGGTGCAAGATTCGCATGTAG 1788
QY 1807 AGCATGTGATCTGTGGGCAAGACCCCGCAAGTTCCGCTGCGCATCAAGAGAGAGCTGTG 1866
DB 1789 AGCATGTGATCTGTGGGCAAGACCCCGCAAGTTCCGCTGCGCATCAAGAGAGAGCTGTG 1848
QY 1867 GAGACTGTGAGAGCGAGCTTACGTGAGAGGCGACCTGTGATCCCGAGTGGAGTTGTGAGAC 1926
DB 1849 GAGACTGTGAGAGCGAGCTTACGTGAGAGGCGACCTGTGATCCCGAGTGGAGTTGTGAGAC 1908

```
OY 1927 ACCCCCCCTGTGAAGCTGTGTACACAGCTGAGAAAGAGCCCATCATCGGCGCGAG 1986
XX |||||
PS 1909 ACCCCCCCTGTGAAGCTGTGTACACAGCTGAGAAAGAGCCCATCATCGGCGCGAG 1968
XX |||||
OY 1987 ACCCTTCTAGTGAAGCGCGCGCCCAACCGCAGAGACCAAGATCGGACCGGCTTACGTT 2046
XX |||||
DB 1969 ACCCTTCTAGTGAAGCGCGCGCCCAACCGCAGAGACCAAGATCGGACCGGCTTACGTT 2028
OY 2047 ACCGACCGGCGCGCGCGCAAGAAATCGTGAAGCTTGAACCAACCAACCAAGACCGAG 2106
XX |||||
DB 2029 ACCGACCGGCGCGCGCGCAAGAAATCGTGAAGCTTGAACCAACCAACCAAGACCGAG 2088
OY 2107 CTGACAGGCGCATCAGCTGAGCCCTGTGACAGACAGCGGAGAGGTGAACATCGTGAACGAG 2166
DB 2089 CTGACAGGCGCATCAGCTGAGCCCTGTGACAGACAGCGGAGAGGTGAACATCGTGAACGAG 2148
OY 2167 AGCCAGTACGCGCTGTGAGCATCATCCAGGCGCCAGCCCGCAAGAGCGAGAGCTGTGT 2226
DB 2149 AGCCAGTACGCGCTGTGAGCATCATCCAGGCGCCAGCCCGCAAGAGCGAGAGCTGTGT 2208
OY 2227 AACCAATATCTGAGACAGCTGTATCAAGAAAGAAAGGTGTACTTGAAGCTGTGTCCCGCC 2286
DB 2209 AACCAATATCTGAGACAGCTGTATCAAGAAAGAAAGGTGTACTTGAAGCTGTGTCCCGCC 2268
OY 2287 CACAAAGGCGATCGCGCGCGCAACGAGCAGATCGAACAGCTGTGAGCAAGGCGCATCCGCAAG 2346
DB 2269 CACAAAGGCGATCGCGCGCGCAACGAGCAGATCGAACAGCTGTGAGCAAGGCGCATCCGCAAG 2328
OY 2347 GTGCTGTCTTGAAGCGGCGCATCGATGCGGCGCATCTGTATCTACAGTACATGAGCAAGCTG 2406
DB 2329 GTGCTGTCTTGAAGCGGCGCATCGATGCGGCGCATCTGTATCTACAGTACATGAGCAAGCTG 2388
OY 2407 TACGTGGGAGAGGCGCGCGCTTGAAGTTCATTAAGCTTCCCGGCGCTTGAACCGGCT 2463
DB 2389 TACGTGGGAGAGGCGCGCGCTTGAAGTTCATTAAGCTTCCCGGCGCTTGAACCGGCT 2445
```

RESULT 13
ADCI3230
ID ADCI3230 standard; DNA; 3930 BP.

```
XX AC ADCI3230;  
XX 18-DEC-2003 (first entry)  
XX DE DNA of HIV construct GagComp1Polmut_C SEQ ID NO 9.  
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;  
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.  
XX OS Human immunodeficiency virus.  
XX PN WO2003004620-A2.  
XX PD 16-JAN-2003.  
XX PF 05-JUL-2002; 2002WO-US021420.  
XX PR 05-JUL-2001; 2001US-0303192P.  
XX PR 31-AUG-2001; 2001US-0316860P.  
XX PR 16-JAN-2002; 2002US-0349871P.  
XX PA (CHIR ) CHIRON CORP.  
XX PA (UIST-) UNIV STELLENBOSCH.  
XX PI Zur Megele J, Barnett SW, Llan Y, Engelbrecht S, Van Renburg RJ,  
XX DR MPI; 2003-221593/21.  
XX PT New expression cassette comprising a polynucleotide sequence encoding a  
XX PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,  
XX PT Prot, or Rev polypeptide, useful for immunization, or generating
```

PT packaging cell lines.

XX Disclosure; Fig 6; 301pp; English.

XX The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 97.0%; Score 2394.8; DB 10; Length 3930;

Best Local Similarity 99.2%; Pred. No. 1.2e-288;

Matches 2431; Conservative 0; Mismatches 7; Indels 12; Gaps 2;

```
OY 14 TGGCGAGGCGCATGAGCGGCGCACGCGCCCAACATCTGTATGACGCGGCAACTTCA 73
DB 1487 TGGCGAGGCGCATGAGCGGCGCACGCGCCCAACATCTGTATGACGCGGCAACTTCA 1546
OY 74 AGGCGCCCAAGCGCATCATTAAGTCTTCAATGCGGCGCAAGGAGGCGCATCGCGCGCA 133
DB 1547 AGGCGCCCAAGCGCATCATTAAGTCTTCAATGCGGCGCAAGGAGGCGCATCGCGCGCA 1606
OY 134 ACTGCGGCGCGCGCGCGCAAGAGGCGCTGTAAGTGTGCGCAAGAGGCGCACAGATGA 193
DB 1607 ACTGCGGCGCGCGCGCGCAAGAGGCGCTGTAAGTGTGCGCAAGAGGCGCACAGATGA 1666
OY 194 AGAAGTGAACCGGAGCGCGCAAGGCACTTCTTCCGAGAGGAGCTGTGCTTCCCGCAAGGCA 253
DB 1667 AGAAGTGAACCGGAGCGCGCAAGGCACTTCTTCCGAGAGGAGCTGTGCTTCCCGCAAGGCA 1726
OY 254 AGGCGCGGAGGTTCCCGCGGCGGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 313
DB 1727 AGGCGCGGAGGTTCCCGCGGCGGCGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1786
OY 314 AGGTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 373
DB 1787 AGGTGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1846
OY 374 TCCCGCGAGATCACCTGTGTGAGGCGCGCGCTGTGTGAGATCAAGGTGTGCGGCGGCGGATCA 433
DB 1847 TCCCGCGAGATCACCTGTGTGAGGCGCGCGCGCTGTGTGAGATCAAGGTGTGCGGCGGCGG 1906
OY 434 AGGAGGCGCTGTGTGAGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 493
DB 1907 AGGAGGCGCTGTGTGAGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1966
OY 494 GCAAGTGAAGCGGCAAGATGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 553
DB 1967 GCAAGTGAAGCGGCAAGATGATCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2026
OY 554 ACCAGATCTGTATGAGATCTGTGCGGCGAAGAGGCGCATGTGCGGCGGCGGCGGCGGCGG 613
DB 2027 ACCAGATCTGTATGAGATCTGTGCGGCGAAGAGGCGCATGTGCGGCGGCGGCGGCGGCGG 2086
OY 614 CCCCCGTGAACATCATTCGGCGCGGCAATGCTTCAACCCAGCTGTGCGGCGGCGGCGGCGG 673
DB 2087 CCCCCGTGAACATCATTCGGCGCGGCAATGCTTCAACCCAGCTGTGCGGCGGCGGCGGCGG 2146
OY 674 CCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 733
DB 2147 CCATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2206
OY 734 TGAAGCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 793
DB 2207 TGAAGCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2266
```


Db	3215	TGATCTGGGGCAAGACCCCCAAAGTTCCGCCCTGACCATTCCAGAAAGAGACCTGGGAGACT	3334
Oy	1874	GGTGACCCGACTACTGTGAGGCCACCTTGGAATCCCGAGTGGAGTTGTTGTAACACCCCC	1933
Db	3335	GGTGAACCGACTACTGCGGACGGCCACTTGAAATCCCGAGTGGAGTTGTTGTAACACCCCC	3394
Oy	1934	CCCTGGTGAACCTGTGTGTACCAAGCTGGAGAGAGGCCCATCATCGGGCCCGAGACTTTCT	1993
Db	3395	CCCTGTGTGAACCTGTGTGTACCAAGCTGGAGAGAGGCCCATCATCGGCCCGAGACTTTCT	3454
Oy	1994	ACGTGGAACGCGCGCCGCCAACCCGCGAGACCAAGATCGGCAAGCCGCGCTACGTGACCGAAC	2053
Db	3455	ACGTGGAACGCGCGCCGCCAACCCGCGAGACCAAGATCGGCAAGCCGCGCTACGTGACCGAAC	3514
Oy	2054	GGGGCCGGCAGAAAGATCGTGAAGCTGACCGAGACCAACCAACAGAAAGCCGAGCTGACAG	2113
Db	3515	GGGGCCGGCAGAAAGATCGTGAAGCTGACCGAGACCAACCAAGAAAGCCGAGCTGACAG	3574
Oy	2114	CCATTCAGCTGGCCCTTGAGAGACAGCGGACGAGGTGAACATGCTGACCCGACAGCCAGT	2173
Db	3575	CCATTCAGCTGGCCCTTGAGAGACAGCGGACGAGGTGAACATGCTGACCCGACAGCCAGT	3634
Oy	2174	AAGCCCTGGGGCATCATCCAGGCCCAAGCCCGACAAGAGGAGAGAGGAGCTGGTGAACCAAG	2233
Db	3635	AAGCCCTGGGGCATCATCCAGGCCCAAGCCCGACAAGAGGAGAGGAGGAGCTGGTGAACCAAG	3694
Oy	2234	TCATTCGACAGCTGATTCAGAAAGAGAGAGGTGTACTTGAGCTGGGTGCCCCCAAGG	2293
Db	3695	TCATTCGACAGCTGATTCAGAAAGAGAGAGGTGTACTTGAGCTGGGTGCCCCCAAGG	3754
Oy	2294	GCATTCGGGGGCAACAGAGCATCCGACAGCTGTGAGGAAGGGCATTCGCAAGGTGCTGT	2353
Db	3755	GCATTCGGGGGCAACAGAGCATCCGACAGCTGTGAGGAAGGGCATTCGCAAGGTGCTGT	3814
Oy	2354	TCCTTGAACGGGATCGATGGCGGCATTCGTGATCTACCACTAATGAGACGACTGTGACGTGG	2413
Db	3815	TCCTTGAACGGGATCGATGGCGGCATTCGTGATCTACCACTAATGAGACGACTGTGACGTGG	3874
Oy	2414	GCAAGGGCGGGCCCTTAGATTCGATTTAAAGCTTCCCGGGGCTTAGCACCGGT	2463
Db	3875	GCAAGGGCGGGCCCTTAGATTCGATTTAAAGCTTCCCGGGGCTTAGCACCGGT	3924

XX	WP1: 2003-221593/21.
XX	New expression cassette comprising a polynucleotide sequence encoding a
XX	polypeptide including an HIV Gag, Env, Int, Nef, p15maeH, Pol, Tat,
XX	Pro, or Rev polyprotein, useful for immunization, or generating
XX	packaging cell lines.
XX	
XX	Disclosure: Fig 8; 301pp; English.
XX	
XX	The invention relates to a novel expression cassette comprising a
XX	polynucleotide sequence encoding a polyprotein including an HIV Gag, Env,
XX	Int, Nef, p15maeH, Pol, Tat, Pro, or Rev polyprotein. The novel
XX	expression cassette can be used to treat HIV type C by gene therapy or
XX	used in the development of a vaccine. The gene delivery vector is
XX	administered intramuscularly, intradermally, intravenously,
XX	subcutaneously, intradermally, intradermally, intravenously,
XX	intracranially, orally or intravenously. The expression cassette is useful
XX	for immunisation, generating packaging cell lines and producing HIV
XX	polyproteins. This polynucleotide sequence represents the DNA of an HIV
XX	Type C related sequence of the invention.
XX	
XX	Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;
XX	
XX	Query Match 96.9%; Score 2393.2; DB 10; Length 3930;
XX	Best Local Similarity 99.2%; Pred. No. 1.9e-288;
XX	Matches 2430; Conservative 0; Mismatches 8; Indels 12; Gaps 2;
XX	
XX	14 TGGCGGAGGCGCATGAGCGAGGCGCAACGCGCAATCTGATGACGCGAGCAATTCA 73
XX	1487 TCGCGGAGGCGCATGAGCGAGGCGCAACGCGCAATCTGATGACGCGAGCAATTCA 1546
XX	
XX	74 AGGCGCGCGAGCGCATCATCAAGTGTCTTCAACTGCGGCGAGGAGGCGCAATCGCGCGCA 133
XX	1547 AGGCGCGCGAGCGCATCATCAAGTGTCTTCAACTGCGGCGAGGAGGCGCAATCGCGCGCA 1606
XX	
XX	134 ACTGCGCGCGCGCGCGCGCAAGAGGCGCTGCTGGAAGTGGCGGAGAGAGGCGCAAGATGA 193
XX	1607 ACTGCGCGCGCGCGCGCGCAAGAGGCGCTGCTGGAAGTGGCGGAGAGAGGCGCAAGATGA 1666
XX	
XX	194 AGGATGTGACCGAGGCGCGAGGCGCAACTCTTCTCGGAGAGACTTGAGCTTCCCGAGGCA 253
XX	1667 AGGATGTGACCGAGGCGCGAGGCGCAACTCTTCTCGGAGAGACTTGAGCTTCCCGAGGCA 1728
XX	
XX	254 AGGCGCGCGAGTTCCTCCGAGCGAGAGAGAGCGCGCGCAAGGCGCGCAAGCGGAGCTGC 313
XX	1727 AGGCGCGCGAGTTCCTCCGAGCGAGAGAGAGCGCGCGCAAGGCGCGCAAGCGGAGCTGC 1786
XX	
XX	314 AGGTGCGCGCGAGCAACCCCGCGAGCGCGCGCGCGCGCGAGCGCGCAAGGCGCAACTGAACT 373
XX	1787 AGGTGCGCGCGAGCAACCCCGCGAGCGCGCGCGCGCGCGAGCGCGCAAGGCGCAACTGAACT 1846
XX	
XX	374 TCCCGCGAGTCAACCTGTGAGAGAGCGCGCGCGCGCGCGCGAGTGAAGTGGCGCGAGATCA 433
XX	1847 TCCCGCGAGTCAACCTGTGAGAGAGCGCGCGCGCGCGCGCGAGTGAAGTGGCGCGAGATCA 1906
XX	
XX	434 AGGAGCGCGTGTGAGCAACCGCGCGCGAGCAACCGGTGTGAGAGAGATGAAGCTTGCCTG 493
XX	1907 AGGAGCGCGTGTGAGCAACCGCGCGCGAGCAACCGGTGTGAGAGAGATGAAGCTTGCCTG 1966
XX	
XX	494 GCAAGTGAAGCGCAAGATGATCGCGCGCATCGGCGCTTCAATCAAGTGGCGCGAGTACG 553
XX	1967 GCAAGTGAAGCGCAAGATGATCGGCGCGCATCGGCGCTTCAATCAAGTGGCGCGAGTACG 2028
XX	
XX	554 ACCAGATCTGATGAGATCTGCGCGCAAGAGGCGCATGCGGACCGGTGTGATTCGCGCGCA 613
XX	2027 ACCAGATCTGATGAGATCTGCGCGCAAGAGGCGCATGCGGACCGGTGTGATTCGCGCGCA 2086
XX	
XX	614 CCGCGGTGAACATCATCGCGCGCAACATGTGAGACCGAGCTGGGCTGCAACCTGTAATTCTC 673
XX	2087 CCGCGGTGAACATCATCGCGCGCAACATGTGAGACCGAGCTGGGCTGCAACCTGTAATTCTC 2146
XX	
XX	674 CCATGACCGCGCATGAGACCGTGCCTGCGGAGCTGAGACCGCGGATGAGACGCGCGCAAGG 733

Db 2147 CCATGAGCCCTCGAGACCGTGGCCGTGAAGCTGAAGCCGGCATGAGCGCCCAAG 2206
Qy 734 TGAACAGTGGCCCTGACCGAGAGAGAGATGAAGCCCTGACCCGCATCTGAGAGAGA 793
Db 2207 TGAACAGTGGCCCTGACCGAGAGAGAGATGAAGCCCTGACCCGCATCTGAGAGAGA 2266
Qy 794 TGAAGAGAGAGAGAGATCAACCAAGATCGGCCCGAGAACCCCTTACAACACCCCGTGT 853
Db 2267 TGAAGAGAGAGAGAGATCAACCAAGATCGGCCCGAGAACCCCTTACAACACCCCGTGT 2326
Qy 854 TCGGCATCAAG 913
Db 2327 TCGGCATCAAG 2386
Qy 914 ACAAGCGACCGAGAGCTTCTGAGAGAGTGCAGACTGGGCAATCCCAACCCCGCGCGTGA 973
Db 2387 ACAAGCGACCGAGAGCTTCTGAGAGAGTGCAGACTGGGCAATCCCAACCCCGCGCGTGA 2446
Qy 974 AGAAG 1033
Db 2447 AGAAG 2506
Qy 1034 AG 1093
Db 2507 AG 2566
Qy 1094 GCATCGGCTACAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1153
Db 2567 GCATCGGCTACAGAGTCAAG 2626
Qy 1154 AG 1213
Db 2627 AG 2686
Qy 1214 ACCAGTACATGAGACGACTGTACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1273
Db 2687 ACCA-----GGCCCCCTGTACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2740
Qy 1274 AGATGAG 1333
Db 2741 AGATGAG 2800
Qy 1334 ACCAG 1393
Db 2801 ACCAG 2854
Qy 1394 TGCAGCCCATGAG 1453
Db 2855 TGCAGCCCATGAG 2914
Qy 1454 TGGGGAG 1513
Db 2915 TGGGGAG 2974
Qy 1514 AGCTGTGCGCGAG 1573
Db 2975 AGCTGTGCGCGAG 3034
Qy 1574 TGGAGCTGGCCGAG 1633
Db 3035 TGGAGCTGGCCGAG 3094
Qy 1634 CCAAGCAAG 1693
Db 3095 CCAAGCAAG 3154
Qy 1694 TCTACCAAG 1753
Db 3155 TCTACCAAG 3214
Qy 1754 ACAACCAAG 1813

Db 3215 ACAACCAAG 3274
Qy 1814 TGAATCTGGGGAG 1873
Db 3275 TGAATCTGGGGAG 3334
Qy 1874 GGTGAG 1933
Db 3335 GGTGAG 3394
Qy 1934 CCTGTGTAAAGTGTGTGTACAG 1993
Db 3395 CCTGTGTAAAGTGTGTGTACAG 3454
Qy 1994 AGGTGAG 2053
Db 3455 AGGTGAG 3514
Qy 2054 GGGGCGGAG 2113
Db 3515 GGGGCGGAG 3574
Qy 2114 CCATTCAG 2173
Db 3575 CCATTCAG 3634
Qy 2174 AGGCGCTGGAGAGATCAAG 2233
Db 3635 AGGCGCTGGAGAGATCAAG 3694
Qy 2234 TCATGAG 2293
Db 3695 TCATGAG 3754
Qy 2294 GCATGCGGCGAG 2353
Db 3755 GCATGCGGCGAG 3814
Qy 2354 TCTGTGAG 2413
Db 3815 TCTGTGAG 3874
Qy 2414 GCAGGCGGCGAG 2463
Db 3875 GCAGGCGGCGAG 3924

Search completed: December 30, 2005, 08:56:53
Job time : 1308.83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8718.85 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-30

Perfect score: 2469

Sequence: 1 gtgcagccaccatgcccga.....gggctagcaccgggtgaattc 2469

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_hic:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_esc7:*
9: gb_gsa1:*
10: gb_gsa2:*
11: gb_gsa3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	136.4	5.5	330	10	CL293849 02S0349-0
2	97.8	4.0	2031	10	CL974899 OaIFCC042
3	97.2	3.9	2886	10	CL967785 OaIFCC015
4	91.4	3.7	1509	10	CL959255 OaIFCC002
5	90.8	3.7	2598	4	AY103647 Zea mays
6	89.4	3.6	743	10	CZ247380 A1AA-aaf3
7	86.6	3.5	1398	10	CL961989 OaIFCC006
8	85.6	3.5	951	3	BM321451 rockefell
9	85	3.4	869	7	CK159167 FGA504056
10	84	3.4	3069	10	CL973991 OaIFCC025
11	83.8	3.4	1941	10	CL971508 OaIFCC021
12	82.6	3.3	1781	8	CK099623 RECM0260
c 13	80.8	3.3	1060	10	CM922203 OaIFCC005
14	80.6	3.3	892	10	CZ216254 A1AA-aaf2
15	80.2	3.2	1132	3	BM320864 rockefell
16	79.6	3.2	1165	3	BM320900 rockefell
17	79	3.2	867	3	BM321430 rockefell
18	78.4	3.2	11691	10	CL962901 OaIFCC008
19	78.2	3.2	1962	10	CL961326 OaIFCC005
20	78.2	3.2	2853	10	CL974397 OaIFCC025
21	78	3.2	1485	10	CL970981 OaIFCC020
22	76.6	3.1	1550	3	BM321022 rockefell

23	76.6	3.1	2559	10	CL982027 OaIFCC046
24	76.2	3.1	1401	10	CL962721 OaIFCC038
25	75.6	3.1	671	6	CA093222 SCCCL300
26	75.4	3.1	2299	4	AY106831 Zea mays
27	75.2	3.0	854	7	CK777127 964131 MA
28	75	3.0	862	3	BM321023 rockefell
29	75	3.0	3822	10	CL972913 OaIFCC023
30	74.6	3.0	914	9	BZ568300 pac82-164
31	74.6	3.0	2010	10	CL962831 OaIFCC008
32	74.6	3.0	2313	10	CL982362 OaIFCC047
33	74.6	3.0	2682	10	CL969033 OaIFCC017
34	74.4	3.0	3134	4	AY109500 Zea mays
35	74.2	3.0	1680	10	CL982770 OaIFCC049
36	74	3.0	1290	10	CL972679 OaIFCC023
37	74	3.0	1386	11	D0045165 Homo sapi
38	74	3.0	2072	4	CR603312 full-1eng
c 39	73.8	3.0	889	7	CK159613 OaIFCC017
c 40	73.6	3.0	757	9	CC678788 OGM171TV
41	73.6	3.0	2028	10	CL979437 OaIFCC003
42	73.2	3.0	853	3	BM321393 rockefell
43	73	3.0	1689	10	CL972373 OaIFCC022
44	72.8	2.9	566	3	BM587428 170006873
45	72.8	2.9	1575	10	CL979927 OaIFCC045

ALIGNMENTS

RESULT 1
LOCUS CL293849/c 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MutTail Library Zea mays genomic clone
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Bukariyola; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 330)
REFERENCE Lathaw,S., Tan,B.C., Settles,A.M. and McCarty,D.R.
AUTHORS Sequence tagged transposon insertions from the UniformMu maize population
TITLE Unpublished (2003)
JOURNAL Contact: Donald R. McCarty
COMMENT Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line: 02S0349-08, Primer, set: A
Class: transposon insertion site.
Location/Qualifiers
1..330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/clone_1ib="UniformMu MutTail Library"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sephadex 400 spin columns and cloned into the TOPO PCR4 vector."

FEATURES

source

ORIGIN

Query Match	5 %;	Score 136.4;	DB 10;	Length 330;
Best Local Similarity	66.4%;	Pred. No.2.8e-14;		
Matches 211;	Conservative	0;	Mismatches 106;	Indels 1; Gaps 1
Qy	421	GGCGGCGAGATCAAGAGAGGCCCTCTGTCGACACCGCGCGCGACGACCGCGTGTGAGAG	480	
Db	329	GGGGGGGCGAGCTGGAGGAAGCTTATTAGATACAGAGCGAGATGATACAGTATTAAAGAA	270	
Qy	481	ATGAGCCTGCGCCCGGCAAGTGGAGGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG	540	
Db	269	ATGACTTTGACAGGAAGATGGAAACCAAAAATGATAGGGGGGAATTGGAGGTTTATCAAA	210	
Qy	541	GTGGCGCCAGTACGACCAATCCTGATTCAGATCTCGCGCAAAAGGCCATCGGCACCGTG	600	
Db	209	GTAAACAGTATGATTCAGGTACCCATAGAAATCTGTGGGCATTAAGCTATAGGTACGGTA	150	
Qy	601	CTGATTCGGAGCCCAACCCCGTGAAATCATCTCGGCGCCCAATCTGTGACCCAGCTGGAGCTGC	660	
Db	149	TTAGTAGAGAACCTACACCTGTCAACTTAATTGGAAGAAATCTGTGACTCAAGATTGG-TGC	91	
Qy	661	ACCCTGAACCTTCCCATCATGACCCCATCGAACCGGTGCCCGTGAAGCTGAAGCCCGGCATG	720	
Db	90	ACCTTAATTTTCCCATTAGTCTCTATTGAAACTGTACCGATTAATTAAAGCAGGAATG	31	
Qy	721	GACGGCCCCCAAGGTGAG	738	
Db	30	GATGGCCCAAAAGTAAAG	13	

RESULT 2	
LOCUS	CL974989
DEFINITION	CL974989 2011 bp DNA linear GSS 21-SEP-2004 OslFCC042695 Oryza sativa Express Library Oryza sativa (Indica cultivar-group) genomic, genomic survey sequence.

ACCESSION	CL974989
VERSION	CL974989.1
KEYWORDS	GI:52404497
SOURCE	GS.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

REFERENCE
1 (Bases 1 to 2031)

Bakayote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Erihartoideae; Oryzaceae; Oryza.

AUTHORS Ma, L., Wang, J., Chen, C., Liu, X., Su, N. H., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Bao, J.

TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis

JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China

Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

```

FEATURES
source
Location/Qualifiers
1..2031
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Express Library"
/note="Oryza sativa exon trapped genomic sequences"

```

Query Match	4.0%	Score 97.8;	DB 10;	Length 2031;
Best Local Similarity	42.6%;	Pred. No. 2.7e-07;		
Matches 787; Conservative	0;	Mismatches 1032;	Indels 30;	Gaps 4

QY 576 CGGCAAGAGGCGCATTCGCGACCGTGTGATCGGCCCCACCCCGGTGAACATCATTCGGCCG 635
 |||||

Db	153	CGGCACGACGTACTCGTCGCTCGAAGTCTACCGGAAACGGACCATGTCTGAGATCATCGCCAA	212
OY	636	CAACATGCTGACCCGAGCTGGGCTGACCCCTGAATCTTCCCATCAAGCCCCATCGAAGCCGT	695
Db	213	CGACCAAGGGAGAACCGGATTCAGCCCTCGTGGGTTCGCTTCAACCGACGCGCGGAGCGGCT	272
OY	696	GCCCGTGAAGTGAAGCCCGGCAATGACGAGCCCAAGGTGAAGCAGTGGCCCTTGAACGA	755
Db	273	CATCGGTGAGGCCGCGCAAGAACACAGGCGGCGGCAACCCGAGCGGAGCCATTCACGACGC	332
OY	756	GAGAGAGATCAAGGCGCTTGAACGCGCATCTCGGAGAGATGAGAAAGAGAGGCAAGATCAC	815
Db	333	CAAGCGGCTCATCGGCGCGGAGTTCTCGGACGCGGAGGTGACGCGCATGATGAAGTGTCT	392
OY	816	CAAGATCGGCCCCCGAAGAACCCCTTACAAACACCCCGCTTCCGATTCAGAAAGAAAGACAG	875
Db	393	GCCGTTCCCGTCGTCGACCGGAACGAGCAAGCCGACCGCTCGAGTGAAGACGCG	452
OY	876	CACCAAGTGGGCGCAAGCTGTGTGAATTCCGCGAGCTGAAACAAAGCCACCCAGAGCTTTG	935
Db	453	CGACGTGG-----CGTGTTCAGCCCGGAGGAGGTGAACGCGCCATGATGTCTCAC	500
OY	936	GAGGTGAGCTGGGCAATCCCCACACCCGCGCGCTGAAAGAAAGAAAGAGGCGTACCGT	995
Db	501	GCGATGAGAGACGCGCCGAGGCTTACTCGGCAAGAGGTCAACCGCGCGCGTCTCAC	560
OY	996	GCTGACGTGGGCGACGCGCTTACTTTCAGCGTGCCTCTGACGAGGACTTCCGACAGTACAC	1051
Db	561	CGTCCCGGCTTACTTTCACACGACGCGACGCGACGCGACGACGACGACGCGCGTCTATCGC	620

OY	1056	CGCCTTCACATATCCCGAGATCAACAAAGAGACCCCGGCTATCGGTACCAAGTACAAAGT	1115
Db	621	CGGGCTCACCTTCACACGGCATCATCAACAGACCCACCGCCCGCCATATGGCTTACGGCAT	680
OY	1116	GCTGCCCCAGGGCTGGAAGGGGCAACCCAGATCTTTCAGAGGAGCATGACCAAGATCTCT	1175
Db	681	CGACAAAGAGGGGCGCGAAGAAAGATCTGTCTTTCGACTCGCGGGGAGAGTTTGA	740
OY	1176	GGAGCCCTTCCGCGCGCCCGCAACCCCGAGATCGTATATCAACAGTACATGACGACCTTGA	1235
Db	741	CGTACAGATCTCTCGCATTCGACAAAGCGGTGTTGAAGTCTCTTGCACCAAGCGGACAC	800
OY	1236	CGTGGGACAGCGACCTTGAGATCGGCGAGACACCGGCAAGATTCAGAGAGCTGCGCAGCA	1295
Db	801	CCACCTCGCGCGGCAAGGACTTTCGACCAACGCTTCATGCAACATTCGTCAAGGTCAATCG	860
OY	1296	CCTGCTGCGCTGCGGGCTTTCACCAACCCCGCAACAAGAGCAGAGAGAGCCCTTCTCT	1355
Db	861	CCGGAAGCACGGGCGCGGACATCAACGGGAGACGGCGCGCGCTGTGGGCAAGCTTCGCGCGGA	920
OY	1356	GTGATGAGCTTACGAGCTGCAACCCCGCAAGATGACCGTGCAAGCCCATCGAGCT---GCC	1412
Db	921	GTGCGAGCGGCCCAAGCGCGCGCTTCAGACAAACAGACCAAGGTGCGCTGCGAGTTCGAGTTC	980
OY	1413	CGAGAAGGAGAGCTGGAACCGTGAACGACATCCAGAAAGCTGTGTGGGCAAGCTGAACCTGAGGC	1472
Db	981	CTTGTTCGACGGCGCTCGACTTGTGGAGCCGCTCTCCCGGCGCGGTTTCGAGAGACTCAA	1046
OY	1473	CAGCCAGATCTTACCCCGGACATCAAGGTGCGCAGACTGTGCAAGCTGTCTGCGCGGCCAA	1533
Db	1041	CAGCGACTCTTTCAGAAAGACGATGGTGTCGGTGAAGAAAGCCATATGCGGAGCGCCGGCT	1100
OY	1533	GGCCTTCACCGACATCTGTGCCCCCTGACCCGAGAGAGCGCGTGAAGCTGTGCGAGAACCG	1592
Db	1101	GAGCAAGGGCGACATTCGACGAGATCTGTTCTGTGCGCGGCGACACCAAGATTCGCCAAGT	1166
OY	1593	CGAGATCTCTGCGCAGGCCCGCTGACACGGCGTGTACTACGACCCCAAGCAAGACCTTGTGTGC	1652
Db	1161	GCAGAGAGTGTCTCAAGAGCTACTTTCGTGTGGCAAGAGGCCCAACCGCGCGCTCAACCCCGA	1222
OY	1653	CGAGATTCAGAGCAAGGGCGACGACCAAGTGAGCTTACCAAGTCTTACCAAGAGCCCTTCAA	1712
Db	1221	CGAGGCGGTGCGCTGATACGGCGCGCGCGTGCAGAGCCAGCATATTCAGCGCGCACGTTGACGA	1288

QY 1713 GAACCTGAAGACCGGCAAGTAGCCCAAGATGCGCACCGCCCAACCAACGAGCTGAAGCA 1772
 |||||
 DB 1281 GAACACGAGAGCATATGATCTCTCGACCTCGCGCCGCTCACCTTCGCGCTGAGACGCGC 1340
 QY 1773 GGTGAACGAGGCGCGTGCAGAAATGCGCANTGGAAGACATGATCTGGGGCAAGACCCC 1832
 |||||
 DB 1341 CGGCGCGGTGATGAGCCAGATCTATCCCGGCAACACGCTGTGCGCAGCAAGAGAGCGCA 1400
 QY 1833 CAAGTCCCGCTCCCATCCATCCAGAAAGAGACCTTGAGAGACCTGTGTGACCGACTACTGAGCA 1892
 |||||
 DB 1401 GGTGTTTCAACACGTACAGAGACAGAGACACACGTCGACATTCAGAGTGTGAGAGGCGCA 1460
 QY 1893 GGCACCTGTGATCCCGAGTGGAGATTCTGTAACACCCCCCTGTGTGAAGCTGTGTGA 1952
 |||||
 DB 1461 GCGGAGCATAGACAGAGAGCAACCGGCTGTGCTCGCAGATTCGACCTTCGCGCATCCGCGC 1520
 QY 1953 CCAAGCTGAGAAAGAGGCCATATCGGCGCCGACCTTCTATGATGTGAACGCGCGCCGCA 2012
 |||||
 DB 1521 GCGCGCGAGGAGGCGCGCCGCAAGTGAAGTGAAG-----TTGAGAGTGAACGCGCA 1571
 QY 2013 CCGGAGACCAAGATCGGCAAGGCGCGCTACGTCAGACCGGAGCGCGCAAGAGATCGT 2072
 |||||
 DB 1572 CCGCATCTCTACAGCTGTCTGCGCGCCGCAAGGCGCACCGGAGGTCCGAGAGATCACTAT 1631
 QY 2073 GAGCTGACCGAGACCAACCAAGAGACCGAGCTGCAAGGCGCATCACTGTGCGCTGCA 2132
 |||||
 DB 1632 CTCGCGGACGACCGCAAGATCGCAGAGAGAGATCGACCGGATGTGTGCGCAGAGCGCA 1691
 QY 2133 GGAAGCGGCGAGAGAGTGAATCTGTGACCGACACCGATGAGCCCTGGGCTATCTCA 2192
 |||||
 DB 1692 GGAAGTTGCGCGAGAGAGACCGCGGCAACAGGAGAGAGTGAAGCGCCGCAACGCTGGA 1751
 QY 2193 GSCCGACCGGCAAGAGAGAGAGAGCTGTGTAACAGAT-----CATCGAGACGT 2246
 |||||
 DB 1752 GCGCTACGTCTACATCAATCAAGACACGCTCGCGCGCAGATGCGGAGCGCATGAGAGG 1811
 QY 2247 GATCAAGAGAGAGAGTGTACTGTAGCTGAGTGCCTCCGCGCAAGAGGCGATCGGCGCA 2306
 |||||
 DB 1812 CGAGAGAGAGAGACAGTGTGAGAGGCGGTGAGGAGAGCGTACGATGTGCTGAGACGCA 1871
 QY 2307 CGAGCAGATCGACAGCTGTGTGAGCAAGGCGCATCCGAGAGTCTGTCTGTGACGCGCAT 2366
 |||||
 DB 1872 CCGGAGCGCGCGAGAGAGATGACAGAGAGCTGAGAGACTGAGAGACCTGTGCA 1931
 QY 2367 CGATGCGGCGATGTGATCTACAGATACATGACGACCTGTACGTGGGC 2415
 |||||
 DB 1932 CCGCTCATGTGCGCGGTCTACAGAGAGTCCGCGCGCGCGCGCGC 1980

RESULT 3
 CL967755 2886 bp DNA linear GSS 21-SEP-2004
 LOCUS OslfPC015718 Oryza sativa Expressed Library Oryza sativa (indica
 DEFINITION cultivar-group) genomic, genomic survey sequence.
 ACCESSION CL967755
 VERSION CL967755.1 GI:52390149
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Bkharotyca; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriharotidae; Oryzaceae; Oryza.
 1 (bases 1 to 2886)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W., and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 CONTACT Chen Chen
 DEPARTMENT Department of Bioinformatic
 BEIJING Institute of Genomics

Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-8048676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Location/Qualifiers
 source 1..2886
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN

Query Match 3.9%; Score 97.2; DB 10; Length 2886;
 Best Local Similarity 43.2%; Pred. No. 3.5e-07;
 Matches 712; Conservative 0; Mismatches 903; Indels 33; Gaps 4;

QY 189 GATGAAGACTGCACCGAGCGGCAAGCTTCTTCCGAGAGACTTGCTTCCCA 248
 |||||
 DB 219 GGTGATGAGGCGCGCGCCATATGCGCTCTGTCATCACTCACTCACTT 278
 QY 249 GGGCAAGCGCGGAGTTCCGAGGAGCAAAACGCGGCAACAGCCCAACCGCGCA 308
 |||||
 DB 279 CCACTGAGAGACTTCTGCGCATGCTGCTCTCTGTCATCACTCACTT 338
 QY 309 GCTGAGGTGCGC---GGCAGAACCCCGCAGCGAGCGCGCGCGCGCGCGCGCAGC 365
 |||||
 DB 339 CATGAGAGAGAAATATGCGGAGAGCGCGCGCGCGCTCATGAGCGCGCTGCGCGCA 398
 QY 366 CCTGAATCTTCCCGAGATACCTGTGAGAGCGCGCGCTGTGAGATCAAGTGGCGG 425
 |||||
 DB 399 GACCAAGGTGCTCAGAGAGAGGAAATGAGAGAGAGAGAGAGCGGTCTCTCTCCCG 458
 |||||
 QY 426 CCAAGATCAAGAGAGCGCTGTGAGACCGGCGCGAGACACCGTGTGAGAGATGAG 485
 |||||
 DB 459 CCAATCATATCAGATCAATCACTGCGGATCACTCCCGAGAGCGCGCTGTCCAGAG 518
 |||||
 QY 486 CCGCGCGCAAGTGAAGACCAAGATGATGCGCGCATGCGCGCTTATCAAGGTGCG 545
 |||||
 DB 519 CACCGCGCTCAAGTGTGAGCAAGCGCGCGCTCAACCGCGAGTGTGATCCGTCACAAAGCA 578
 |||||
 QY 546 CCAAGTCAACAGATCTGATGAGATCTGCGGCAAGAGAGCGCATGCGCACCGTGTAT 605
 |||||
 DB 579 CCGCGGAGAGGCGGTCTTCTCGGAGTCAACCGTCAAGAGAGCGGAGATGAGAGCGGTG 638
 |||||
 QY 606 CCGCGCGCAACCGGTGAATCATATGCGCGCAAGATGATGAGATGAGATGAGATGAG 665
 |||||
 DB 639 CATGCGCACCGGAGTCACTTCTTGGCAAGAGCGCGCATCTGTGTGACAGACCA 698
 |||||
 QY 666 GAACTTCCCATATGAGCCCATGAGACGATGAGCGCGCGTGAAGCTGAAGCCCGCATGAGCG 725
 |||||
 DB 699 CAACATGCGGCACTTCACTGCTGTGCTCAACGCGCATGCGCAATTTGTGATCATTCAT 758
 |||||
 QY 726 CCGCAAGTGAAGCACTGCGCGTGAACGAGAGAGATCAAGAGCGCTTGAACCGCATCTG 785
 |||||
 DB 759 CCGCGTGCATGATATATGAGATCATCTGTATGATCCGATCAACACCGCGCTGACCG 818
 |||||
 QY 786 CCAAGAGATGAGAGAGAGGCAAGTCAACAGATGCGCGCGGAGAACCCCTTCAACAC 845
 |||||
 DB 819 CAGCGCATATGAGAACCTGCTGCTGCTCATGCGCGGATCCCATGCGCATGCCAC 878
 |||||
 QY 846 CCGCGTGTGCGCATCAAGAGAGAGACAGACCAAGTGGCGAGCTGTGATCTTCCG 905
 |||||
 DB 879 CCGTGTGTGCGTACATGAGCATGCGGTGCGTCCACCGCGGTGTGCAAGAGCGCGCATAC 938
 |||||
 QY 906 CAGCTGAACAGCGACCCAGACTTCTGAGAGTGTGAGCTGAGGATCTCCCAACCCCGC 965
 |||||
 DB 939 CAAGCGAGATGACCGCATGAGAGAGTGGCGGCGCATGAGTGTCTGCGACCAAGAG 998
 |||||
 QY 966 CCGCTGAAGAGAGAGAGAGCGTGAACGCTGAGTGTGAGCGTGTGAGCGCTTCAAGCGT 1025
 |||||

```

Db      999  CGGAGCGCTGACGCTCAACAGCTCAACCTCCAGCAAGAGCTGATGAGGTGTACGGGCG 1058
Oy      1026  GCCCTGAGCAGAGACTTCGCGAAGTACACGGCTTTCAACATCCCGAGCATCAAGAGA 1085
Db      1059  GGGGCTGAGCAAGAGACTCGGTCTCTGTTACGGGGGAGGGGCTCCCGGTGAGAGACCA 1118
Oy      1086  GACCCCGGCGATCCGCTACCAAGTACACGTGTGCCCCAGGGCTGAGAGGGAGCCCGAG 1145
Db      1119  GAGCGCCATTGACACATGATGATGTGGGACATGTCCCGGAGCCCGAGAGAGCCCGGGCG 1178
Oy      1146  CATCTTCCAGAGACGATACCAAGATCTTGGAGCCCTTCCGCGCCCGAGACCCCGAGAT 1205
Db      1179  CAT-----CAAGAGAGTCACTTCTCCCGTTCAACCCCGTGCAGAGACGCGACGCG 1229
Oy      1206  CGTGAATCTACAGTACATGAGAGCACTGTACGTTAGGGCAGAGCCTGAGATCGGCGAGCA 1265
Db      1230  CATCACTTACATGACGCGGCAATGGCAATGGACAGAGATCAAGAGAGGGCGCGCGAGCA 1289
Oy      1266  CGCGCCCAAGATCGAGAGCTGCGCAAGACCTGTGCGCTGGGGCTTCAACACCCCGCA 1325
Db      1290  GATCATCGAGCTGTGCAAGATGAGCAAGAGAGCGCCGAGAGAGAGGTGCAACGCTGATGA 1349
Oy      1326  CAAGAAAGCACCAAGAGAGAGCCCTCTTCTGTGATGGCTTACAGAGCTGCAACCCGAGCA 1385
Db      1350  CCAAGTACGCGAGACCGCGGCTCGGTGCTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 1409
Oy      1386  GTGAGACCGTGCAGCCCATGAGCTGCGCGAGAGAGAGAGCTGAGACCGTGTGAAGCATCCA 1445
Db      1410  G-----AGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1454
Oy      1446  GAAGCTGTGGGCAAGCTGAACTGAGGCGCAGCGAGATCTAACCCCGCATCAAGGTGGCGCA 1505
Db      1455  GCGGCTGTTCGACCGCGCGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1514
Oy      1506  GCTGTGCAAGCTGCTGCGCGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1565
Db      1515  CGGCGGTGACGTGAAGATGATACCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1568
Oy      1566  GGCAGAGCTGAGAGCTGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1625
Db      1569  GGGCGGCGGCTCGGAGATGGGACCAAGATGTAACCGGTGAGAGAGAGAGAGAGAGAGAGAG 1628
Oy      1626  CTACGACCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1685
Db      1629  GAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1688
Oy      1686  CTACCAAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
Db      1689  CGGCGGTGTTCCCGAGAGACAGAGTACAGAGTGTGAAGCGCTTCAGAGAGAGAGAGAGAG 1748
Oy      1746  CACCGCCCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Db      1749  CTGCGGATGACCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1808
Oy      1806  GAGCATGTGATCTGTGGGCAAGACCCCGC 1833
Db      1809  CATCGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1836

```

```

RESULT 4
LOCUS      CL959255
DEFINITION OricCC002520 Oryza sativa Expressed Library Oryza sativa (indica
ACCESSION  CL959255
VERSION     CL959255.1 GI:52373273
KEYWORDS    GSS.
SOURCE      Oryza sativa (indica cultivar-group)
ORGANISM    Oryza sativa (indica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE   1 (bases 1 to 1509)

```

```

AUTHORS    Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
            Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yan, L.,
            Wang, X., Deng, X. W. and Wang, J.
            An analysis of transcriptional regulation of the rice genome and
            its comparison to Arabidopsis
JOURNAL     Unpublished (2004)
COMMENT     Contact: Chen Chen
            Department of Bioinformatics
            Beijing Institute of Genomics
            Chinese Academy of Sciences, Beijing 101300, China
            Tel: 86-10-80481559
            Fax: 86-10-80488676
            Email: chenchen@genomics.org.cn
            Rice genomic sequence.
            Class: exon-trapped.
FEATURES    source
            location/Qualifiers
            1..1509
            /organism="Oryza sativa (indica cultivar-group)"
            /mol_type="genomic DNA"
            /db_xref="taxon:3946"
            /clone_lib="Oryza sativa Expressed Library"
            /note="Oryza sativa exon trapped genomic sequences"
ORIGIN
Query Match      3.7%; Score 91.4; DB 10; Length 1509;
Best Local Similarity 43.8%; Pred. No. 3.7e-06;
Matches 395; Conservative 0; Mismatches 506; Indels 0; Gaps 0;
Oy      430  ATCAAGAGAGCCCTGCTGAGACCGGCGCGAGAGACACCGTGTCTGAGAGAGATGAGCTTG 489
Db      1  ATGAGCAGATCATATCGGGCCATCTTCGCTGACATCCCGGATCGGTCAACCCCGCTCG 60
Oy      490  CCGGCAAGTGAAGACCCCAAGATGATCGGCGGCGATCGGCGCTTATCAAGGTGGCGAG 549
Db      61  CCGGCGGCTGCTCTTCGCTGCTTGTGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Oy      550  TACGACGAGATCTGATGAGATCTGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
Db      121  GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Oy      610  CCCACCCCGTGAACATCATGAGCGCGCAACATGCTGACCCAGCTGAGCTGACCTTAAAC 669
Db      181  CGCTCCCGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Oy      670  TTCCCATACGCGCCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 729
Db      241  TGGGCTTCACGCGCTGCTGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
Oy      730  AAGTGAAGCAGTGGCCCTTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 789
Db      301  CAAATCTTGTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Oy      790  GAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 849
Db      361  CGGTTCCGGTGGTCCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Oy      850  GTTTCGCTATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 909
Db      421  GCGCTCTGCGGTGCTGCGGCTCTTCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Oy      910  CTGAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 969
Db      481  TTCCTCAAGCTGAGCGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Oy      970  CTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1029
Db      541  CTCGCGCGCTTACTTGGCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Oy      1030  CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
Db      601  GCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Oy      1090  CCGGCAATCGGTACAGTACAGTGTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149

```

Db 661 AAGAGCTGACCGCTGCTCCAGTGCAGCCCGCGCTGGAAGAGCTCGCTTCGTC 720

Qy 1150 TTCGAGAGCAGATGACCAAGATCTCGAGAGCCCTTCGCGCCGCAACCCGAGATCGTG 1209

Db 721 ACCTACCAACAGCCAGACGTCGCGCTCCGCGAGCGAGCTTCGCTGCGCTCTTC 780

Qy 1210 ATCTACAGTACATGACGACCTGTACGTGGGAGGAGACTGTGATTCGCGCCAGACCCG 1269

Db 781 TGGATGTTCATCGCGGAGAGAGCTCGCGCTGTGTCGCGCCGCGCTCGAGCGGCTCATC 840

Qy 1270 GCCAAGATGAGAGAGCTGGCGAGAGCAGCTGTGCGGTGGGAGCTTCACACCCCGACAG 1329

Db 841 CTGTGAGACGGTGGCTTCGACCGCGCTGTCCGCTGATTTCCGACACAGGATCAAG 900

Qy 1330 A 1330

Db 901 A 901

RESULT 5
AY103647 2598 bp mRNA linear HTC 18-FEB-2005

LOCUS Zea mays PCO142084 mRNA sequence.
DEFINITION AY103647
ACCESSION AY103647.1 GI:21206725
VERSION
KEYWORDS
SOURCE Zea mays
ORGANISM Zea mays
Bukariyora; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 2598)
Gardiner, J., Schroeder, S., Polacco, M.L., Sanchez-Villada, H., Fang, Z., Morgan, M., Landewe, T., Fessler, K., Uesche, F., Hanfey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C. and Coe, B.H. Jr.
Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional plant hybridization
Plant Physiol. 134 (4), 1317-1326 (2004)

JOURNAL PUBMED 15020742

REFERENCE 2 (bases 1 to 2598)
Hailey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whiteitt, M.S., Arthur, L.W., Hanfey, M., Morgan, M. and Tingey, S.V.
Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
Unpublished (2002)
3 (bases 1 to 2598)
Coe, B.H.

JOURNAL REFERENCE
TITLE Direct Submision
AUTHORS Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA

COMMENT If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schabber, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.

FEATURES
source
Location/Qualifiers
1..2598
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeGDB:638378"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/Dupont Consensus Library"
/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize

ORIGIN Mapping Project"

Query Match 3.7%; Score 90.8; DB 4; Length 2598;
Best Local Similarity 43.3%; Pred. No. 4.9e-06;
Matches 719; Conservative 0; Mismatches 909; Indels 33; Gaps 5;

Qy 170 GCGGCAAGAGGCGCCACAGATGAGAGCTGACCCGAGGCGCAGGCACTTCTCCGG 229

Db 351 GAGGAGAGTGAATCTCGATCGCATCGCCAGCGGCGGACATCTAGCGGTACACAG 410

Qy 230 AGACCTGAGCTTCCCGGAGGAGGAGGCGGAGGATTCGCCAGGAGCAACCGGCGCA 289

Db 411 GCTTGGGAGGACCTTCACCGCGGACCAAGAGCGGCGCGCTCCAGGTGACCTCG 470

Qy 290 ACAGCCCAACAGCGGAGCTGCAAGTGGCGGCGCAACCCCGCAGCGAGCGCGG 349

Db 471 TCAGGCACTCAACCGCGGATCTTCGCAACCGGCGAGCGAGCGGCAACGCTCGG 530

Qy 350 CCGAGCGGAGGAGCACTTGAATCTTCCCGAGATCACTCTGAGGAGCGGCGGCTGTA 409

Db 531 AGGTACCTCGCGGAGATGTGTGCGCATCAACCTCTTCCAGGCTACTCCGCGCA 590

Qy 410 GCATCAAGGTGGGCGGCGAGATCAAGAGGCGCTGTGACACCGGCGCGGACGACCG 469

Db 591 TCCGCTTGAAGATCTTCAGAGGCTTCAGAGAGTGTCTCAACCGGTTCAGCCCTGCG 650

Qy 470 TGTGAGAGAGATGAGCTTCCCGGAGGTGAAGGCGCAAGATGATCGGCGGATCGCG 529

Db 651 TCGCGCTCGGGGAGCATCAACCGGATCGGGGAGCTGTCCGCTCTCATATCGCGG 710

Qy 530 GCTTCATCAAGTGGGCGGAGTACAGACAGATCTGTATCGAGATTCGCGGCAAGAGCGCA 589

Db 711 GCTTCATCAAGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 767

Qy 590 TCGGAGCGGTGTGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 649

Db 768 AGCGCGCGGAGGCGGCTTCAAGATCGCGGATGAGAGGCGGCTTTCAGAGTCAACCGCA 827

Qy 650 AGCTGGGCTGACACCTGAATCTTCCCGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 709

Db 828 AGGAGGCGCTCGCATCGTCAACGCGAGCGTCCGCGGCTCGCGGCGGCGGCGGCGGCGG 887

Qy 710 AGCGCGGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 769

Db 888 TGTACGACCGCAACGCTCTGCGCGCTCTGTGCGAGGTCTGTCTCGCGGCTTCTGCGAGG 947

Qy 770 CCGTACCGGCGATGTGCGGAGATGAGAGAGGCGGCAAGATCACCAAGATCGGCGCGG 829

Db 948 TGTGAGCGGAGAGCGGCGGAGTACGAGCACTTGAACCAAGCTGAGAGCACACCGG 1007

Qy 830 AGAACCTTACCAACACCGCGGCTGTTCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 889

Db 1008 GGTTCATGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1067

Qy 890 AGCTGTGAGCTTCCGCGAGCTGAGACAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 949

Db 1068 AGCGCAAGAGTGAAGAGAGCTGAGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1115

Qy 950 GCATCCCGCAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1009

Db 1116 ACAGGTAGCGGCTCGCGAGTGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1175

Qy 1010 AGCGCTTACTTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1069

Db 1176 CGCGCAACAGTCAATCGAGCGGAGGTCACTCGGTGAACAGCAACCGGCTATCGAGG 1235

Qy 1070 CCAAGATCAACAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1129

Db 1236 TCACCGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1295

Qy 1130 GGAAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1189

Db 1236 TGGACAGCGCCGCTCGCATCGCCACATCGGCAAGCTCATGTTCCGCGACTTCTCCG 1355
 QY 1130 CCCGCAACCCGAGATCGATCTACCACTAGATGAGAGACTGTATCGTGGGCAAGGA-- 1247
 Db 1356 AGCTGTCAACAGATTCTACCAACACAGGCTCACTTCAACTGGGCGGCAAGCCGCAAC 1415
 QY 1248 ----CCTGAGATCGGCGAGCAACCGCGCCAGATCGAGAGGCTGCGCAAGCACTGTCTC 1303
 Db 1416 CCAAGCTGAGCTACCGGCTTCAAGGGGCAACGAGATCGCANTGGCCTCTACTGTCTCCAGC 1475
 QY 1304 GCTGGGGCTTCAACCAACCCCGCAAGAGACCAAGAGAGGCCCTTCTGTGTGATGG 1363
 Db 1476 TCACATGCTCTGGGCAACCCCATCAACCAAGTGAAGCGCGGAGAGCAACCAAG 1535
 QY 1364 GCTACAGCTGACCCCGCAAGTGAACGCTGAGAGCCCATGAGCTGCGCGCAAGAGAGA 1423
 Db 1536 AGTGAATCTCTTGGGCTCTGTCTGAGGAGAGAGCCCGCAAGGATGACATCTGA 1595
 QY 1424 GCTGAGCCGTGAACGACATCGAAGAGCTGTGGCAAGCTGAATGGGCAAGCAAGATCT 1483
 Db 1596 AGCTAGTGTCTCACTCATCATGTGTGGGCTGTGGCAGCGCGTGAACCTGGCGCACTCG 1655
 QY 1484 ACCCGGATCAAGATGTCGCGCAAGCTGTGCAAGCTGTGCGCGCGCAAGCCCTGACCG 1543
 Db 1656 AGAGAAACATCAAGGCTGTGTGAAGAACACCGTGAACCGAGTGGCCAGAGAGTGTGA 1715
 QY 1544 ACATCGTGTCCCTGACCGAGAGAGCCGAGCTGGAAGTGGCGAGAA-----CGCGC 1594
 Db 1716 CATGAACCTCTCGGCGAGGCTCTCAAGCGCCGCTTCAAGCAAGAGAGCTGATGAGG 1775
 QY 1595 AGATCTGTGCGAGCCCGTGAACGCGGTGATCAAGACCCAGCAAGACCTGTGTGGCG 1654
 Db 1776 CATGACACCGGAGCGCGTGTTCAGGTACGCGAGAGACCGCGGCGAGCGCTGCGCG 1835
 QY 1655 AGATCAAGAGAGAGGCGCGCAAGCAAGTGAACCTAACAATCTAACAGAGCCCTTCAAGA 1714
 Db 1836 TGATCGAAGAGTGGCGCGCGTGTGTGTGAACCAACCCCTCAAGCAGCG---GCGAGCGCG 1892
 QY 1715 ACCTGAAGACCGGCAAGTACGCGCAAGATGCGCACCGCGCAACCAAGAGTGAAGAGCG 1774
 Db 1893 AGCGGAGCGCTTCCGTGTCTCAAGATCAACGAGTTCAAGAGAGAGCTTCCGCGCGTGC 1952
 QY 1775 TGACGAGAGCGGTGCAAGATGCGCATGAGAGATCTGTG 1815
 Db 1953 TGCCCCAGAGGTGAGAGCGCGCGGTGTGGCGTGCAG 1993

RESULT 6
 C2247380 743 bp DNA linear GSS 11-FEB-2005
 LOCUS A1A4-asf35a12.b1 Ancylostoma caninum whole genome shotgun library
 DEFINITION (A1A4GSS 001) Ancylostoma caninum genomic, genomic survey sequence.
 ACCESSION C2247380
 VERSION C2247380.1 GI:59632821
 KEYWORDS GSS.
 SOURCE Ancylostoma caninum (dog hookworm)
 ORGANISM Ancylostoma caninum
 Bacteria; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatidae; Ancylostoma.
 1 (bases 1 to 743)
 Miteva,M., McCarter,J.P., Pape,D., Ritzer,E., Tsagaris,I., R.,
 Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messing,D.,
 Waterson,R.H., Clifton,S.W. and Wilson,R.
 Genome Survey sequences from the parasitic nematode Ancylostoma
 caninum
 Unpublished (2004)
 CONTACT: Miteva M
 Washington University in St. Louis
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: nematode@watson.wustl.edu

Genomic DNA provided by John Hawdon (mtmj@wumc.edu) DNA
 sequenced by Washington University Genome Sequencing Center
 Class: shotgun.
 Location/Qualifiers
 1..743
 /organism="Ancylostoma caninum"
 /mol_type="genomic DNA"
 /strain="Baltimore"
 /db_xref="taxon:29170"
 /dev_stage="Adult"
 /lab_host="GS10"
 /clone_lib="Ancylostoma caninum whole genome shotgun
 library (A1A4GSS 001)"
 /note="Vector: pOTM13; Site 1: BstXI; Site 2: BstXI;
 Ancylostoma caninum genomic DNA was randomly sheared,
 end-repaired and size fractionated to enrich for 2-4 kb
 fragments. Genomic DNA was provided by John Hawdon
 (mtmj@wumc.edu) at George Washington University.
 Sequencing by Washington University Genome Sequencing
 Center, St. Louis, MO."

Query Match 3.6%; Score 89.4; DB 10; Length 743;
 Best Local Similarity 46.3%; Pred. No. 8e-06;
 Matches 294; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY 751 ACCGAGGAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGCGCAAG 810
 Db 63 AACGAC 122
 QY 811 ATACCAAGATGCGCGCGAGAACCTTACACACACCGGTGTGCGCATCAAGAGAG 870
 Db 123 AAC 182
 QY 871 GACAGACCAAGTGGCGGAGAGTGTGTGCTTCCGAGCTGAGCAAGCGCACCCAGAG 930
 Db 183 AAC 242
 QY 931 TTCTGGAGGTGCACTGTGGCATTCCTCCACCGCGCGGTGTAAGAGAGAGCGTG 990
 Db 243 GAC 302
 QY 991 ACCGTGTGAGAGTGGCGGACGCTTACTTCAAGCTGCGCTTGGAGAGAGCTTCCGCAAG 1050
 Db 303 AAC 362
 QY 1051 TAGACCGCTTCAACATCCCGACATCAACACAGAGAGACCCCGGATCTTCAACATAC 1110
 Db 363 AACGAC 422
 QY 1111 AACGTGTGCGGCGGAGAGGCGAGCGCGACGATCTTCCAGAGCAGATGACCAAG 1170
 Db 423 AAC 482
 QY 1171 ATCTGAGAGCTTCCGCGCGCGACACCCGAGATGTGATCTTACAGTACATGAGACAC 1230
 Db 483 AAC 542
 QY 1231 CTGTAGTGGGCGAGCCTGAGATCGGCGACACCGCGCAAGATGAGAGAGCTGCGCG 1290
 Db 543 AACGAC 602
 QY 1291 AAGCAGCTGTGCGGTGGGGCTTCAACACCCCGCAAGAGACCAAGAGAGAGCGCCCG 1350
 Db 603 AAC 662
 QY 1351 TTCTGTGAGTGGCTACAGAGCTGACCCCGACAA 1385
 Db 663 TAC 697

RESULT 7
 C1961989

LOCUS	CU961989	1398 bp	DNA	linear	GENS 21-SEP-2004
DEFINITION	Os16FCC00691 Oryza sativa Expressed Sequence Library Sequence.				
ACCESSION	CU961989				
VERSION	CU961989.1	GI:52378720			
KEYWORDS	GSF.				
SOURCE	Oryza sativa (indica cultivar-group)				
ORGANISM	Oryza sativa (indica cultivar-group)				
REFERENCE	Bukarjota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.				
AUTHORS	1 (bases 1 to 1398) Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiong, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.				
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis				
JOURNAL	Unpublished (2004)				
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.				
FEATURES	Location/Qualifiers				
source	1..1398 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Expressed Library" /note="Oryza sativa exon trapped genomic sequences "				
ORIGIN					
Query Match	3.5%;	Score 86.6;	DB 10;	Length 1398;	
Best Local Similarity	45.9%;	Pred. No. 2.7e-05;			
Matches	508;	Conservative 0;	Mismatches 569;	Indels 30;	Gaps 5;
Oy	1	GTCCAGCGCACCATGATGCGGAGGCGCATGAGCGACGACGAGCGCCAAACATCTGATGAC	60		
Db	166	GTTCAGAGGCACTGTGGGAGGCGCCACATCTGTTCTGTTCTCGTAAACACCCCAACAG	225		
Oy	61	CGCAGCACTTTCAGAGGCGCCCAAGCGCATCATCAATGATCTTCACTGCGGCAAGAGGAC	120		
Db	226	GCCCGGCGCTCGGGCGCGGCAAGGCGCGCG---ACCTACCTATCGGAGAGCGCGGCG	282		
Oy	121	CACATCGCCCGCACTGCGCGCGCGCGCGCGCGCAAGAGGCGTGTGGAAGTGCAGAGAG	180		
Db	283	CGAATGATTCGCGCGGTGCGCGCATTCGACCAAGAGTCTGTGAGAAATCCACCTGTGCG	342		
Oy	181	GGCCACCAAGTAAAGACTGCACCGAGCGCCAGGCGCACTTCTTCGCGAGGACCTGTGCG	240		
Db	343	GTCAAGACCGCGAGGCGCATTCAGAGAGTCTTCGACCAACAGCGCGCGGCGCTGTGCG	402		
Oy	241	TTCCCCAGGAGGAGCGCGCGAGTTTCCCAAGCGAGCAAGAACCGCGCAACGCGCCAC	300		
Db	403	TTTCAGATCTCTCCACCCCGAGTTCTCTCGCGCAAGGCGCGCCATTCGCGACCTGTCTC	462		
Oy	301	AGCGCGAGCTGCAAGTGTGCGCGCGCAACCCCGCGAGGAGGCGCGCGCGCGCGCGAG	360		
Db	463	GCCCCCGA-----CGCGTCTCTCATTCGCGCGCGCGGAGACCGCGCGGCGCGCGCG	513		
Oy	361	GGCACCCTGAACTTCCCCCAAGTACCTCTGTGGACGCGCGCGCTGTGAGATTCAGAGT	420		
Db	514	GCGTTCAGGCGCGCTCAAGACGTGTACGCGCGCATGCGTCTCCGAGAGAGATC---CTC	570		
Oy	421	GGCGGCGAGATCAAGAGGCGCTGTGTCGACACCGGCGCGAGCAACCGTGTGTGAGAG	480		
Db	571	ACCAACCACTTGTGTGCGCGCGAGCTGTTCAGAGCTCGCGCGCAAGCGCTTCTGTGCG	630		
Oy	481	ATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGGCGATCGCGGCTTCATCAAG	540		

Db	Accession	Version	Source	Definition	Location/Qualifiers
Db	631		AGGATCTCGTCCGTCAACGCGCATGTCGCGCTCGGAGGCGCACCGCGCGCAAGTCGCC	690	
Qy	541		GTGCGCGCAGTACGACCAAGATCTGATCGA---GATCTGCGGCGAAGAGCCATCTGGCAC	597	
Db	691		GAGGTGCGCTTACCGCGGGCGAAGACTCCAGATCGGCCCGCAAGTTCTTCACGCCAGC	750	
Qy	598		GTGCTATCGGCGCCCAACCCCGGAAACATCATCGGCGCGCAACATGCTGACCGACTGGGC	657	
Db	751		GTGGAGATTCCGTGGCTCTCTCTTCCAGAAAGACATCTTCAACTCTGCTG-----	798	
Qy	658		TGCACCTGAACTTCCCATCAAGCCCATCGAGACCGTGCCTGTAAGCTGAACTGAAACCCGAC	717	
Db	799		TACATCTGCGAGTGCATATGGCTCTCCCGAGGTGGCGCAACTAGTGAAGCAGGTTATCAAG	858	
Qy	718		ATGACGCGGCCCAAGGTGAAGAGTGGCCCCCTTACGAGAGAAATCAAGGCCCTGACC	777	
Db	859		ATCAACGACTACCAAGAAAGACCGGTTCTGTGAACCGCGTGGTGTCTCATGTTCAACAC	918	
Qy	778		GCCATCTGCGAGAGATGAGAAAGAGGGCAATACCAAGATCGGCCCGCGAAGACCC	837	
Db	919		GTGCGCGCGCAAGAAAGATCGCGGTGCTGGGGTCTCGCTTCAAGAAAGACACCGCGCAC	978	
Qy	838		TACAAACACCCCGGTTTCCGCATCAAGAAAGAGACGACCAAGTGGCGCACTGGTG	897	
Db	979		AGGAGAGCGCCCGCATCGATCGTTCGAGAGGGCTGATCGGCCCAAGGCCAAGGTGAC	1038	
Qy	898		GACTTCCGCGAGTGAACAAAGCGCACCCAGACCTTCTGGAGGTGACCTGGGCATCCC	957	
Db	1039		ATCTAGACACCCCGAGTGAAGAGAACCAAGTCCAGCGCGACCTCGCATGAGCAAGTTC	1098	
Qy	958		CACCCCGCGGCTGAAAGAAAGAGAGCGTGAACGCTGACCGTGGCGACCGCTTAC	1017	
Db	1099		GATTGGATGATCCCGGTCTCATCTCCAGCCGATGAGCCCAACGGCATCAAGAGTGTCC	1158	
Qy	1018		TTCAAGCGTGGCCCTGAGCGAGGACCTTCGCAAGTACACCGGCTTCAACATCCGAGATC	1077	
Db	1159		GTGCGCTGGAGACGCTTACGAGCGCGGCGGACGCGCACCGCGCTGTGATCTTCAACGAG	1218	
Qy	1078		AACAAAGAGACCCCGCGCATCCGCTTAC	1104	
Db	1219		TGGAGCGAGTTCAAGAGACCTCGACTAC	1245	
RESULT 8	BM321451				
LOCUS	BM321451				
DEFINITION	rockefeller.0.1211 Masticigmoeba balamuthi lambda ZAP II Library				
	Masticigmoeba balamuthi cDNA similar to adenoviral nucleocapsid protein (BC				
	3.3.1.1), mRNA sequence.				
ACCESSION	BM321451				
VERSION	BM321451.1				
KEYWORDS	EST.				
SOURCE	Masticigmoeba balamuthi				
ORGANISM	Masticigmoeba balamuthi				
REFERENCE	Bkaryote; Pelobiontida; Masticigmoebidae; Masticigmoeba.				
AUTHORS	1 (bases 1 to 951)				
	Babteete, E., Brinkmann, H., Lee, D.A., Moore, D.V., Sensen, C.W.,				
	Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and				
	Philippe, H.				
	The analysis of 100 genes supports the grouping of three highly				
	divergent amoebae: Dictyostelium, Entamoeba, and Masticigmoeba				
	Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)				
TITLE	Contact: Muller Mike				
JOURNAL	Laboratory of Biochemical Parasitology				
PUBMED	The Rockefeller University				
	1230 York Avenue, New York, NY 10021, USA				
	Email: mmuller@rockefeller.edu				
	Insert length: 951 Std Error: 0.00				
FEATURES	POLYA=No.				
SOURCE	Location/Qualifiers				
	1..951				


```

/organism="Maestigamoeba balamuthi"
/mol_type="rRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_lib="Maestigamoeba balamuthi lambda ZAP II library"
/notes="syn: Phreatamoeba balamuthi"

```

ORIGIN

Query Match 3.5%; Score 85.6; DB 3; Length 951;
 Best Local Similarity 46.0%; Pred. No. 3.9e-05;
 Matches 323; Conservative 0; Mismatches 376; Indels 3; Gaps 1;

```

1734 GCGCAAGATGCGACCGCCACACCAACGCTGAAGAGCTGACCGAGCCGTCAGAA 1793
1734 GCGCAAGATGCGCTGTGTCTGTGCAACATCTTCTGACGAGAGCAACGCGCGCGC 278
1794 GATGCCATGAGAGCATCTGTATCTGGGGCAAGACCCCAAGTTCCGCTGACATCA 1853
279 CATCCGCGAGGCGCGCTCTCGGCTTCCCTGGAAGGCGGAGAACTTCAAGAGTATG 338
1854 GAAAGAGACTGGGAGACTGTGTGACCGACTGACGAGGCACTGGAATCCCGAGTG 1913
339 GAGAGTCACTGGAAAGCCCTGTGTCTTGGCCCTTACCGAGGACCTCAGATCATGTCA 398
1914 GGAAGTTCGTGAACACCCCTCTGTGAAGCTGTGTATCACTGAGAAAGAACCAT 1973
399 CGAGCGCGGTGACGCGACTGTATGATCAACAAGGGGTTCCGGCCGAGAGCAACCCCA 458
1974 CATCGGCGCGAGACCTTCTATCTGTGACGCGCGCCCAACCGCGAGACCAAGATTGGCA 2033
459 GCTGTGAGAGAGCAAGAGGCGCTCGAGAGGTCTGCTCAACAACGTCTCAAGCA 518
2034 GGGCGGCTTACGTCACCGACCGGGGCGCGAGAAATCGTAGCCCTGACCCGAGACCA 2093
519 GGTTCAGAAAGAGACGACCCCGCTTCTGTGCAAGAAATCTCTCCGAGATCCGCGTGTGAG 578
2094 CCAGAAAGACGAGCTGACAGGCGCATCCAGCTGACCTGCGAGAGACGAGCGAGAGATGAA 2153
579 CGAGGAGAGCAAGCATGGCGTATAGAGGTATACAGAGTGCACCGCGAACGACTGT 638
2154 CATGTCATCCGACAGCGCATGATGACCTGTGGGCAATCATCCAGGCGCCGAGCAAGAGCA 2213
639 GTTCCCGGCGTCAACGTCAACGAC---TCTNTACCAAGAGCAAGTTGCAACAACATCTA 695
2214 GAGCGAGCTGTGAACCAAGATCATGAGAGGAGCTGTATCAAGAGAGAGAGGTACTGAG 2273
696 CGGCTGCGCGCACTGCTCATGAGCGAGCATCAAGGCGGCGACCGAGTATCTCGGCGG 755
2274 CTGGGTGCGCGCGCAAGAGGCGATGCGCGCAACGAGCAGATGCAACAAGCTGTGAGCA 2333
756 CAAGGTGCGCGGTGTGTGCGGGGTAGGGGCAAGTGGGCAAGGGCTGCGCGAGTCTGCG 815
2334 GGGCATTCGCAAGGTGCTGTTCTTGTGACGCGCATGATGAGCGGCGCATGTGATTCACATTA 2393
816 CGGCGAGGCGCTGCGCGTATCATGTGACGAGATGCAACCCATCTGCGGCTCAAGCGCTG 875
2394 CATGACGACCTGTACGTGGGAGGCGGCGCCCTTGAATGCA 2435
876 GATGGCGGCTTCAAGGTCAACACCTCGAGGCGGGGCTCGA 917

```

RESULT 9

CK159167/c

LOCUS CK159167 869 bp mRNA linear BST 05-DEC-2003
 DEFINITION FGAS040564 Triticum aestivum FGAS: Talt5 Triticum aestivum cDNA,

ACCESSION CK159167
 VERSION CK159167.1 GI:38985053

KEYWORDS EST
 SOURCE Triticum aestivum (bread wheat)

ORGANISM

Triticum aestivum (bread wheat);
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE

1 (bases 1 to 869)
 Allard, F., Crosby, W. L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,
 Genswein, B., Graf, R., Gulick, P., Hrycan, L. D., Laroche, A.,
 Links, M. G., McCarthy, E. L., Monroy, A., Muzak, I., Nilsson, D.,
 Peniket, C., Roach, J. L. and Sarhan, F.

AUTHORS

Peniket, C., Roach, J. L. and Sarhan, F.

TITLE

Functional Genomics of Abiotic Stress in Wheat and Canola Crops

JOURNAL

Unpublished (2003)

Contact: Wm L. Crosby

Bioinformatics

University of Saskatchewan, Department of Computer Science

1C101 Engineering Building, 57 Campus Drive, Saskatoon,

Saskatchewan, S7N 5A9, Canada

Tel: 306 966 1769

Fax: 306 966 2033

Email: fgsa_scs@cs.usask.ca

This sequence is the direct result of the Base calling software

Phred (default parameters). It is the raw base calls. To aid in the

identification of the high quality insert the software Lucy

(default parameters) has been run on this sequence. Lucy identified

the region [128,636].

Plate: Talt537 row: N column: 23.

Location/Qualifiers

1..869

/organism="Triticum aestivum"

/mol_type="rRNA"

/cultivar="wheat line PI 178383"

/db_xref="taxon:4565"

/lab_host="DHS alpha"

/clone_lib="Triticum aestivum FGAS: Talt5"

/notes="Organ: Crown; Vector: pGEM-T; SSH (suppression

subtractive hybridization) cDNA library from genotype

PI178383 cold hardened at 2 C for 21 days and 49 days

(equal amount of cDNA pooled together before subtraction,

tester) and subtracted against genotype Norstar cold

hardened at 2 C for 1 day (24 H) (driver). Modified Smart

cDNA (Clontech) priming and non-directional cloning"

ORIGIN

FEATURES

source

Query Match 3.4%; Score 85; DB 7; Length 869;
 Best Local Similarity 45.4%; Pred. No. 5e-05;
 Matches 304; Conservative 0; Mismatches 365; Indels 0; Gaps 0;

ORIGIN

```

580 AAGAGGCGATCGGACCGTGTATCGGCGCCACCGCGTGAATCATCGCGCGCAAC 639
827 ATGGCGGCGCGCGCAACAGATCAACCAACGAGAGCAACCAACCAACCAAC 768
640 ATGTCAGCCGAGCTGGGTGACCTTGAATTTCCCATCAAGCCCATGAGACCGTGGCC 699
767 CACAGAGAGCAACAGGAGCAACAGCAACCAACCAACCAACCAACCAACCAAC 708
700 GTGAAGCTGAAGCCCGCATGAGCGCGCCCAAGGTGAAGAGTGGCCCTTGAACGAGAG 759
707 ACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAAC 648
760 AAGATCAAGGCGCTGACCGGCATCTGCGAGGAGATGAGAGAGGAGGAGATCAAC 819
647 AACACACAGACACACACACACACACACACACACACACACACACACACACACAC 588
820 ATCGCGCGCGAGAACCTTCAACACACCGCGTGTGTGCGCATCAAGAGAGAGAGAG 879
587 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 528
880 AAGTGGCGAGAGCTGTGATCTTCCGCGAGCTGAACAGCGCACCCAGAGCTTGTGGAG 939
527 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 468
940 GTGCGAGTGGGATCCCGACCGCGCGCGCTGAAGAGAGAGAGAGAGAGTGAACG 999
467 AACACACACACACACACACACACACACACACACACACACACACACACACACAC 408
1000 GACGTGGCGAGCGCTTACCTTCAAGCTGCGCTTGAACGAGACCTTCGCAAGTACAC 1059

```

[illegible]

LOCUS	DEFINITION	RESULT 10
CL973991	3069 bp DNA linear	GSS 21-SEP-2004
CL973991	Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.	

ACCESSION	CL973391
VERSION	CL973391.1
KEYWORDS	GI:52402507
SOURCE	GSS.
ORGANISM	Oryza sativa (indica cultivar-group)
	Oryza sativa (indica cultivar-group)

REFERENCE

Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
1 (bases 1 to 3069)

TITLE An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis

JOURNAL COMMENT Unpublished (2004)

CONTACT: Chen Chen

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

```

FEATURES
source
location/Qualifiers
1..3069
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:35946"
/clone_lib="Oryza sativa Express library"
/note="Oryza sativa exon tripped genomic sequences "
```

ORIGIN	Query Match	3.4%	Score 84;	DB 10;	Length 3069;
	Beet Local Similarity	42.8%	Pred. No. 8.2e-05;		
	Matches 536;	Conservative	0;	No. Matches 710;	Indels 6; Gaps 2
Oy	37	ACCAAGCGCAACATCTGTATGACAGCGGCACTTCAAGGGCCCAAGCGCATCATCAAG			96
Db	655	AGCGGCATCTGGGAAGACGACGCTGGCGAGCTACCTTACCGAGGAACAGCATCGGAAG			684
Oy	97	TGCTTCACCTGGCGCAAGAGGGCCATCGCCCGCACTGCGCGCCCCCGCAAGAG			156
Db	685	CACCTTCGCGCTGCACGGCGCTGGCTCCACGCTCCCCCAAGTTCCGCTTCGCCACCTCTTC			744
Oy	157	GGCTCTGGAAGTGGCGGCAAGAGGGCCACAGATGAAAGATGTGACCGAGGCGCAGGC			216
Db	745	CACGACGTATCTGGCGCAGGCCAAGCGCTCTTCGGCGAAGCAAGACTCGCGCGCAGAGC-			803

[illegible]

QY 517 GGGGGCATCGGCGGCTTCATCAAGGTGGCCAGTACACAGATCTGATGAGATCTGC 576
|||
1102 CGCGGAGGGTGTATCGGCCACGGGAGCCGAGGCCGAAGAGTGGACAGCTCAAGAGC 1161
Db 577 GCGATATACGGGATTCGGGATCGGCGGCTTCATCAAGGTGGCCAGTACACAGATCTGATGAGATCTGC 636

57 GGCAGAGAGCCATCGCAGACCGTGCATCGCCCTCCAGGACATATCGGCTCC 658
 1162 CTGTGTGACATCGACGAGAGGGCTCCAACTGCGGTCAACATCGTCATGCTCGCGGA 1221
 637 AACATGTCACCGAGCTGGGCTGCACCTTAACTTCCCATCAGCCCCCATGAGACGCTG 696

Db	1222	CTCTTGTAGTCCAAAGAGAGAGATGAATGGAGGCCGTGATCCATAGCCTTGATGATACG	1281
Dy	697	CCCGTGAAGCTGAAGCCCGCAGTGAAGCGCCCAAGGTGAACATGTGCCCCCTTGAACGAG	756

Db	1282	CCACCCCGCATGCGCGGAGACAGACCCCGCGGAGCGGAGCCGAGAAAGACAGCCTC	1341
Qy	757	GAGAAAGTCAAGGCCCTTGACCGCATTTGCGAGAGATGAGAAAGAGGGCGAAGATCAC	816
Db	1342	GATGATGCGGAGACGGGCGCGCATTCGACGGACGAAACGAAAGAGACGAGAGAGAGAG	1401

[illegible]

OY	937	GAGGTGCAAGCTGGGCATCTCCCCACCCCGCGCGCTGAAAGAAAGAAAGAGCTATCCCTG	996
Db	1522	AAGCTGGTCCGGCTGTGGGTTGCAGAGGGGTTTCATGCAAGCGGAGAAACGGGACAGCGGTG	1581
OY	997	CTGACAGTGGGCGACGCCTACTTCAGAGTGCCCTCGACAGAGCATTCGCCGAATGACAC	1054
Db	1582	GAGGAGCAGCGCCGAGAGATGCTTCAAGAGCTCATCTCCCGGTGCTGTGCAGCTCGTC	1644

Accession	Sequence	Position
Oy	1057 GCCTTCAACATCCCCAGCATCAACAAGAGACCCCCGGCATTCGGTACAGTACAAACGTG	1118
Dd	1642 GAGACGGAACGGCGGGCGGGCGAGAGGTGTCGGCCGTGACGATCTCAACAAGGGTGTGTGAC	1700
Oy	1117 CTGCCCCAGGGCTGGAAGGGGAGGCCCCAGCATCTTCCA--GAGCAGCATGACCAAGATC	1173
Dd	1702 TTCTGTCCAGGCGGAGGCGCGCGAGACCACTTCTCTCCATCTGCAACAGCGGCGCGCGGCG	1761
Oy	1174 CTGGAAGCCCTTCCGGGCGCCCGCAACCCGAGATCGTGAATCTACAGTACATGAGACGACTG	1231
Dd	1762 CTCTCCAAACGGCGCGCCCGCTCGCTCGCGCTTCGGCAACCTTCAGACACCGACTTCGCC	1821
Oy	1234 TACGTGGGACAGCACTTGAGATCGGCGACGACCGCGCCAGATCGAGAGC	1285
Dd	1822 GTCATGTCTGAGGGCGGCCCAAGCTCAACACCTCTCTGTGCGACATTCGCGAGC	1873

RESULT	11
LOCUS	CL971508
DEFINITION	CL971508 Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
ACCESSION	CJ971508
VERSION	CJ971508.1 GI:52397596
KEYWORDS	GSS.
SOURCE	Oryza sativa (indica cultivar-group)
ORGANISM	Oryza sativa (indica cultivar-group) Eurycotla; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae; Eubartoideae; Oryzaceae; Oryza. 1 (bases 1 to 1941)
REFERENCE	Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M., Jiao,Y., Sun.N., Zhang.X., Bao.J., Sun.D., Zhao.H., Yuan.L., Wong.G.K.S., Deng.X.W. and Wang.J. An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis Unpublished (2004)
TITLE	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel.: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.
JOURNAL COMMENT	Location/Qualifiers 1..1941 /organism= "Oryza sativa (indica cultivar-group)" /mol_type= "genomic DNA" /db_xref= "taxon:39946" /clone_1lb= "Oryza sativa Express Library" /note= "Oryza sativa exon trapped genomic sequences"
FEATURES	source
ORIGIN	
Query Match	3.4%; Score 83.8; DB 10; Length 1941;
Best Local Similarity	42.9%; Pred. NO. 8.6e-05;
Matches	769; Conservative 0; Mismatches 992; Indels 30; Gaps 6,
Dn	267 CCCGAGCGAAGAACCCTGGCCCAAGCCCCCAGGCCGAAGTGTGCAGATGCGAGGCGGCA 326
Dn	165 CGCGCCAAAGAACCGTGTCGCATTGAACCCAACAACCGCTTGATGAGGAAGCGTT 224
Dy	327 CAACCCCAGCGAGCGCGCGCGCGAGGCGCACACTGAACTTCCCCCATATCC 386
Dd	225 GATTGTGTAGAGTGTCTTCGACCGCTCCGTGTAGATGATTAAGTAAAGCTATGCGCTTCAA 284
Dy	387 CCTGTGAGAGCGCCCTGTGTGTAGCATCAAGTGTGGCGCGCCAGATCAAGAAGCCCTCT 446
Dd	285 GTGTGTGCTGTGGCCCCGGCGACAACCGCATGATGTCTGTCCAGTACAAAGGGGAGAGAAA 344
Dy	447 GGACACCGGCGCCACGACACCGTGTCTGAGGAATGAGCTGTCCCGCGAAGTGAAGCC 506
Dd	345 GCAGTTTCGCGCGAGAGATCTCTGTCAATGTGTGATCAAGATGAGGAGATGCGGGA 404
Dy	507 CAAGATGATGCGCGGCATGTGGCGGCTTCATCAAGGTGCGCGCATGACGACGATCTGAT 566
Dd	405 GCGCTTACTTGGCTGTGTCAATCAAGAACCGCTGTATACCGTCCGGGCTTACTTAAGA 464
Dy	567 CGAGATCTGCGGCAAGAAAGCCATCGGACCGTGTGTGATGTGGCCCCACCCCGTAGAAT 626
Dd	465 CTGCGAAGAGGACGGCCACAAAGAAAGCGGGCGTCAATGCGCGGCTCAADGTATGCGAT 524
Dy	627 CATCGGCGGCAATGCTGATCCGAGCTGGGCTGCAACCTGAATTCTCCCATCAGCCCAT 686
Dd	525 CATCAACGAGCCCAACGCGCGCGCCATGCGCTTACGAGCTTCGACAAAGAGCGACGAG 584
Dy	687 CGAGACCGTGCCCGTAGAGCTGAAGCCCGGCATGAGACGCGCCCAAGTGAAGCAATGGCC 746
Dd	585 CGCGAGGAAGAACGTGTCTATTTTGAACCTCGCGCGCGGACGTTTTGATGTGTCTCTCT 644

Qy	CGTATCTGGGGCAAGACCCCGCATGTTCCGCTGCCATCCAGAAAGAGACCTGGGAGAC	1812
Db	CAGAGAGAGATCGAGAGACCCATCGATGCGCCCATCAGCTGGCTCGACACCAACAGCT	1713
Qy	CTGGTGGACCGACTACTGCGCAGGCGCACTTGATCCCGAGTGGAGATTCTGTAAACACCC	1872
Db	CGCCGAGGCGCAGCGATTCGAGAGACA---AATGAGAGAGCTGGAGGGCATCTGCACACC	1773
Qy	CCCCCTGTGAAGCTGTGTGTACCAAGCTGAGAGAGAGGCCATCATCGGCGCCGAGACTT	1932
Db	GATCATCGCCCAAGATGTACCAAGGGCGCCGCGGCAGCATGGCCGAGGCGATGACGAGGA	1830
Qy	CTACGTGACGAGCGCGCCGCAACCGCGAGACCAAGATCGCGCAAGCGCGGCTA	1992
Db	CGCGCGGCGGCGGCGAGCGGCGCAGAGCGCCCAAGATCGAGAGAGGTTCACATA	1890

	LOCUS	DEFINITION
RESULT 12 CX099623	CX099623	1781 bp mRNA linear EST 03-JUN-2005 RcCM02560 A normalized whole-life-cycle cDNA library of rice <i>Oryza sativa</i> (indica cultivar-group) cDNA clone EI057014, THREBB09, EI057004, EI062D05, EI027114, EI070G16, EI08 5', mRNA sequence.

ACCESSION	CX099623
VERSION	CX099623.1
KEYWORDS	GI:66912775
SOURCE	EST.
ORGANISM	<i>Oryza sativa</i> (indica cultivar-group)
	<i>Oryza sativa</i> (indica cultivar-group)

REFERENCE
1 (bases 1 to 1781)

TITLE	AUTHORS
Features of the expressed sequences revealed by a large-scale	Zhang, J., Feng, Q., Jin, C., Qiu, D., Zhang, L., Xie, K., Yuan, D., Han, B., Zhang, Q. and Wang, S.

JOURNAL
of
Plant J. 42 (5), 772-780 (2005)

COMMENT

1531885
Contact: Wang S
National Key Laboratory of Crop Genetic Improvement
Hauzhong Agricultural University
Wuhan 430070, China
Tel: 86-27-87282044
Fax: 86-27-87287092
Email: shiyingwang@hotmail.com
Seq primer: 17.

```

FEATURES
SOURCE
Location/Qualifiers
1. 1781
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/strain="indica"
/cultivar="Minghui 63"
/db_xref="taxon:33946"
/clone="E1057014, THREB09, E1057004, E1062D05, E1027114, E1070G16, E108"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_lib="A normalized whole-life-cycle cDNA library of rice"
/notes="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; The library is constructed based on the strategy of saturation hybridization with genomic DNA using rice cultivar Minghui 63. This library consists of cDNA from 15 directionally cloned cDNA libraries constructed with different tissues from 9 developmental stages."

```

	ORIGIN
Query Match	3.3%; Score 82.6; DB 8;
Best Local Similarity	43.2%; Pred. No. 0.00014;
Matches 449; Conservative	0; Mismatches 584; Indels 6; Gaps 1

Oy 231 GGACCTGCGCTTCTCCCGCAGGCGCAAGGGCCCGGAGATTCTCCCGACGACGAAACCGCGCAA 290
 Db 194 GGAGCGCGGTGTGACCGGTGCTCGCGCCAGAGAACCCCGACGAGCAAGGTGGCTTGGCGAGAC 253
 Oy 291 CAGCCCCACAGCCGCGAGCTGACAGTGGCCGGCGCGACACCCCGCAGCGCGAGGCGCGCGC 350
 Db 254 GTGCACCAAGACCAACATGTGTGATGTGTTCGCGGAGATCAACACAAAGCGCACCGTCGA 313
 Oy 351 CGAGCGCGCAGGCGCACCCCTGAACTTCCCCCAGATCAGCCCTGTGGCAGGCGCCCGCTGTGAG 410
 Db 314 CTACGAGAGAGTCTGTCGCGCACACTTGC CGGCGCATGGCTTGTGTTCACACAGCTTCGG 373
 Oy 411 CATCAAGGTGGCGCGCCAGATCAAGAGGCCCTTCTGACAACCGCGCCGACGACACCGT 470
 Db 374 CCTGACGCGCAGCCGCGTCGAGAGGTGTCTGTCAATCATGAGCAGCAGGTTCGCCGACATGCG 433

QY	471	CTGTGAGAGATGTAGCCCTGCCCGGCAAGTGAACCCAGAGATGATTCGGCGGCATTCGGCGG	530
Db	434	GAGGGGTGTGACGGCCACTTTCACCAAGCGCCCCGAGAGATTCGGCGCGCGCGACAGGG	493
QY	511	CTTATCATCAAGTGGCGCCAGTACGACGACGATCTGTATGCAATCTCGCGCAAGAGGCAT	590
Db	494	CCACTGTGTCGGCTACGCGCACCGAGACGCCGCCCGAGCTATAGGCCCTCAGCCCACTGTCT	553

[illegible]

614 CAGGCCGACGGCAAGACCAGGTCACCTCGAGTACCTCAACGAGCCGGCCCATGCT 673

Qy	711	GCCTGGCATGGA	CGGCCCCC	CAAGGTGAAG	AGTGGCCCCCTTGACC	CGAGAAAGATCAAGGC	710
Db	674	CCCGGTCCGGGT	CCACCGCTCT	CATCTTCA	CCAGCA	CGAGAGACCGTCA	CAACGA

Oy CCTGACCGCCATCTGCGAGAGATGAGAGAGAGCGCAAGATCACCMAGATCGGCCCGCA 830
 771
 734 CAAATTTGTTGAC-----GACCTCAAGAGCACTCATCAAGCGGTGATCCCCGACAA 787

759 CACAGATCCCCCCC
 801 GAACCCCTAACACACCCCGGTGTTGCCATCAGAGAGAGGACAGACCAAGTGGCGAA 890
 788 GTACTTCACGAGAGAACCATCTTCACCTCAACCCCTCGGACGCTTGTGTCATCGGCGG 847
 891 GCTGTGACCTTCGCGAGCTGTAACAACGCAACCCAGACTTCTTGAGAGTGACGTGGG 950
 848 GCCCACGCGCAGCGCGGCGCTCACCGGCGGAGATATCATTCAGACACTTAAGGCGGATG 907
 Db

QY	951	CATCCCCCA	CCCCCGCCGCTG	AGAGAAGAAAGAGCTG	ACCGTCTG	AGCTGAGG	1011
Db	908	GGGCGCGGA	CCGCGCGCGCGCTT	CTCCGGCAAGGAC	CCGACCAAGGTG	ACCGGAGCG	967
QY	1011	CGCTACTT	CAGCGTGGCCCTG	GAAGAGACTT	CCGCAAGTAC	CCGCTTCA	1070
Db	968	CGCTACAT	CCGCAAGGCGCGCA	AGAGATG	TGCGCAGCGGCT	CGCCGCGCTG	1022
QY	1071	CAGCAT	CACAACGAGAC	CCCCCGGAT	CCGCTACCAAT	CACATCGT	1133
Db	1028	CATCTG	CAGGTGTGTAGCG	CCATGGGCT	CCCCGAGCGCT	CTCCGTTGTG	1087
QY	1131	GAGGGC	AGCCCCCAG	CATCTT	CCAGAGCAGTGA	CAAGATCT	1190
Db	1088	CTACGG	CACCGGCAG	ATCCCCG	ACAGAGATCT	CTCAAGATCGT	1144
QY	1191	CCGCA	ACCCCGAGAT	CGTAT	TACAGTACAT	TGAGACCT	1253
Db	1148	TTTCAG	CGCCCGGAGAT	GATGAC	CATCACTG	CACTCAAGAGGCGG	1207

QY 1251 GGAGATCGGCACGACCGC 1269
||| ||| ||| |||
Db 1208 CAAGACCGCGCGCTACGGC 1226

```

RESULT 13
LOCUS      CM922203/1
DEFINITION EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA
            Acanthamoeba castellanii genomic clone EDCAR29, genomic survey
            sequence.
ACCESSION  CM922203
VERSION     CM922203.1 GI:60228146
KEYWORDS    SSS.
SOURCE      Acanthamoeba castellanii
            Buxariyola; Acanthamoebidae; Acanthamoeba.
ORGANISM    Acanthamoeba castellanii
            (bases 1 to 1060)
REFERENCE    1
            Anderson, I.J. and Loftus, B.J.
            Gene discovery in the Acanthamoeba castellanii genome
            Unpublished (2004)
AUTHORS      Contact: Iain Anderson
            The Institute for Genomic Research (TIGR, www.tigr.org)
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-795-7949
            Fax: 301-838-0208
JOURNAL      Class: shotgun.
COMMENT
FEATURES
    source          location/Qualifiers
                     1..1060
                     /organism="Acanthamoeba castellanii"
                     /mol_type="genomic DNA"
                     /strain="Neff"
                     /db_xref="taxon:5755"
                     /clone="EDCAR29"
                     /clone_lib="A. castellanii, 6-8 kb library from total
                     genomic DNA"
                     /note="Vector: PHOS2"

ORIGIN
Query Match      3.3%; Score 80.8; DB 10; Length 1060;
Best Local Similarity 43.9%; Pred. No. 0.00029;
Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

261 CGAGTTCCCGACGAGCAGACCGGCGCAACAGCCCGACCGGAGCTGCAAGTGG 320
955 CGTGGCGCGGCGAGCTGCGCACTCAAGACCGCGCGTCAATGCGCTGACATGG 896
321 CGGCGCAACACCCCGGAGGAGCGCGGCGCGGCGCGGCGGCGGCGGCGGCGG 380
895 CGCGCTCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 836
381 GATCACCTCTGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 440
835 CAGGAGGTCAAGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 776
441 CCTGCTGACACCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 500
775 GCTGGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 716
501 GAAGCCCAAGATGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 560
715 CGCGCGCGGCGGAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 656
561 CCGATCGAGATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 620
655 GGAGGAGGAGCGCGCGCTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 596
621 GAACATCATCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 680
595 CGAGGACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 536
681 CCCCATCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 740
535 GCGCATCATGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 476
741 GTGGCCCTCGACCGGAGGAGAGATCAAGGCCCTTGACCGCGCATCTGCGAGGAG 797
475 CCGCTTCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 416

```

```

QY      798 GAAGAGGGCAAGATCACCAAGATGCGCCCGAGAACCCCTTACACACCCCGTGTGCG 857
DB      415 GCAAGTGCATCTCGACGCGCGGAGGTATGACCAAGCTCGAGGCGCGGCTTCAGGCTGGA 356
QY      858 CATCAAGAAAGAGACAGACCAAGTGGCGGAGCTGTGACTTTCGCGAGCTGAACA 917
DB      355 GGTGAGGCTCAACCGCATGAGAGAGAGTAAAGAGATCGAAGGACCAAGAGCA 296
QY      918 GCGACCCAGGACTTCTGGAGGTGACCTGGGACTTCCCGACCCCGCGCTGAAGA 977
DB      295 GCGCATTCAGAGAGATCAAGAGAGAACTGTGCGCATCAAGAGAGATGAGCCCTCA 236
QY      978 GAAGAAGAGCGTGAACCGTGTGAGAGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 1037
DB      235 GCGCGGCTACAGAGAGAGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 176
QY      1038 GCACTTCGCAAGTACACCGGCTTTCACCATCCCGAGCATCAACAGAGACCCCGGCA 1097
DB      175 CGAGGACCTCAAGGCTCAAGGCGGAGCGCGGCGGAGCGCAAGAGACTTCGCGACCGCG 116
QY      1098 CCGCTACCAAGTACACCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1149
DB      115 CGATCTGCGCTACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 64

RESULT 14
LOCUS      CZ216254
DEFINITION A1AA-sat23g02.b1 Ancylostoma caninum whole genome shotgun library
            (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION  CZ216254
VERSION     CZ216254.1 GI:59229909
KEYWORDS    GSS.
SOURCE      Ancylostoma caninum (dog hookworm)
            Ancylostoma caninum
            Buxariyola; Metazoa; Nematoda; Chromadorea; Rhabdidae; Strongylidae;
            Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
REFERENCE    1 (bases 1 to 892)
            Mitreva, M., McCarter, J.P., Pape, D., Rieger, B., Tsagaris, V.I., R.,
            Ronko, I., Martin, J., Wylie, T., Dantze, M., Meyer, R., Messina, D.,
            Waterston, R.H., Clifton, S.W. and Wilson, R.
            Genome Survey sequences from the parasitic nematode Ancylostoma
            caninum
            Unpublished (2004)
JOURNAL      Contact: Mitreva M
            Washington University in St. Louis
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: nematode@wustl.edu
            Genomic DNA provided by John Hawdon (mtmj@wumc.edu) DNA
            sequenced by Washington University Genome Sequencing Center
            Class: shotgun.
FEATURES
    source          location/Qualifiers
                     1..892
                     /organism="Ancylostoma caninum"
                     /mol_type="genomic DNA"
                     /strain="Baltimore"
                     /db_xref="taxon:29170"
                     /dev_stage="Adult"
                     /lab_host="G810"
                     /clone_lib="Ancylostoma caninum whole genome shotgun
                     library (A1AAGSS 001)"
                     /note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
                     Ancylostoma caninum genomic DNA was randomly sheared,
                     end-repaired and size fractionated to enrich for 2-4 kb
                     fragments. Genomic DNA was provided by John Hawdon
                     (mtmj@wumc.edu) at George Washington University.
                     Sequencing by Washington University Genome Sequencing
                     Center, St. Louis, MO."

ORIGIN

```


Tue Jan 10 16:35:49 2006

us-09-610-313b-30.rst

Page 14

Qy	919	CGACCCAGACTTCTGGAGGTGACGTGGGCATCCCCACCCCGCC	966
Db	850	AACCGCGTGGCCAGAAAGAGTCCGCGCTGGGGCTAACCCGACGCCCC	897

Search completed: December 31, 2005, 02:31:20
Job time : 8727.85 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 11955.3 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463
Sequence: 1 ggcagcagccaccatgcgcga.....gggctagcaccggtgaatcc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 segs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenBdb1:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_srs:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_hlg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2463	100.0	2463	6	AX455915 Sequence
2	2442.2	99.2	2463	6	AX455914 Sequence
3	2436.2	98.9	2457	6	AX455916 Sequence
4	2046	83.1	2306	6	BD263704 Improved
5	2046	83.1	2306	6	CQ870574 Sequence
6	2046	83.1	2306	6	AR373387 Sequence
7	2027.6	82.3	9166	6	AX427930 Sequence
8	2025.2	82.2	2312	6	BD263706 Improved
9	2025.2	82.2	2312	6	CQ870576 Sequence
10	2025.2	82.2	2312	6	AR373389 Sequence
11	2019.2	82.0	2300	6	BD263705 Improved
12	2019.2	82.0	2300	6	CQ870575 Sequence
13	2019.2	82.0	2300	6	AR373388 Sequence
14	2019.2	82.0	9788	6	AX427936 Sequence
15	2005.2	81.4	9169	6	AX427931 Sequence
16	2001.8	81.3	9194	6	AX427926 Sequence
17	2000.2	81.2	9194	6	AX427925 Sequence
18	2000.2	81.2	12411	6	AX427927 Sequence

19	1996.8	81.1	9785	6	AX427938 Sequence
20	1993.4	80.9	9167	6	AX427933 Sequence
21	1993.4	80.9	9170	6	AX427928 Sequence
22	1993.4	80.9	9782	6	AX427935 Sequence
23	1993.4	80.9	9783	6	AX427934 Sequence
24	1993.4	80.9	9790	6	CS078943 Sequence
25	1993.4	80.9	9792	6	AX427932 Sequence
26	1993.4	80.9	36066	6	CS078956 Sequence
27	1991.8	80.9	9189	6	AX427921 Sequence
28	1988.8	80.7	4353	11	AF287353 Synthetic
29	1972.2	80.1	3015	6	AX455946 Sequence
30	1955.6	79.4	3009	6	AX455987 Sequence
31	1942.6	78.9	4319	6	BD263639 Improved
32	1942.6	78.9	4319	6	CQ870498 Sequence
33	1942.6	78.9	4319	6	AR373322 Sequence
34	1932	78.4	4352	11	AF287352 Synthetic
35	1932	78.4	4352	11	AY789509 Synthetic
36	1886.6	76.6	2577	6	AX457088 Sequence
37	1879	76.3	4343	6	AX306428 Sequence
38	1878.8	76.3	2305	6	BD263702 Improved
39	1878.8	76.3	2305	6	CQ870572 Sequence
40	1878.8	76.3	2305	6	AR373385 Sequence
41	1875	76.1	4341	6	AX306429 Sequence
42	1865.4	75.7	1977	6	AX455954 Sequence
43	1857.8	75.4	1965	6	AX455952 Sequence
44	1857.2	75.4	7344	6	CS078948 Sequence
45	1852.4	75.2	7897	6	AX427923 Sequence

ALIGNMENTS

RESULT 1
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITL

JOURNAL
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
LOCATION/Qualifiers
1. 2463
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN

Query Match 100.0%; Score 2463; DB 6; Length 2463;
Best Local Similarity 100.0%; Pred. No. 26-202;
Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTGAGGCCCACTGGCCGAGGCGCATGAGCCAGGCCCAAGCCCAATCTCGATTCAG 60
DB 1 GTTGAGGCCCACTGGCCGAGGCGCATGAGCCAGGCCCAAGCCCAATCTCGATTCAG 60
QY 61 CGGAGCACTTCAAGGGGCCCAAGCGCATCATGAGTCTTCAACTGCGGCAAGAGGGC 120
DB 61 CGGAGCACTTCAAGGGGCCCAAGCGCATCATGAGTCTTCAACTGCGGCAAGAGGGC 120
QY 121 CACATGCGCCGCACTGCGCGGCCCGCGCAAGAGGCTGTGGAAGTGCGGCAAGAG 180
DB 121 CACATGCGCCGCACTGCGCGGCCCGCGCAAGAGGCTGTGGAAGTGCGGCAAGAG 180
QY 181 GGCACCAATGAAGACTGACCGAGCGCCAGGCCCACTTCTCCGCGAGACCTGGCC 240

181 GGCACAGATGAAGACTGCAACGAGCGCCAGGCGCACTTCCGCGAGGACCTGGCC 240
241 TTCCCCCAAGGCAAGGCCCGCGAGTTCCCAAGAGACAAACCGCGCCAAAGCCCAAC 300
241 TTCCCCCAAGGCAAGGCCCGCGAGTTCCCAAGAGACAAACCGCGCCAAAGCCCAAC 300
301 ACCCGGAGCTGCAAGGTGCGCGGCGACAAACCCCGGAGAGGCGCGGCGAGCGCGAG 360
301 ACCCGGAGCTGCAAGGTGCGCGGCGACAAACCCCGGAGAGGCGCGGCGAGCGCGAG 360
361 GGCACCTTGAACCTTCCCGAGATCAACCTGTGAGCAGCGCCCTGTGTAGCATCAAGTG 420
361 GGCACCTTGAACCTTCCCGAGATCAACCTGTGAGCAGCGCCCTGTGTAGCATCAAGTG 420
421 GCGCGGCAATCAAGAGAGCCCTGTGTAGCAACCGGCGCGAGCAACCGTGTGTAGAGAG 480
421 GCGCGGCAATCAAGAGAGCCCTGTGTAGCAACCGGCGCGAGCAACCGTGTGTAGAGAG 480
481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGAGCGGCAATCGCGGCGCTTCAATCAAG 540
481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGAGCGGCAATCGCGGCGCTTCAATCAAG 540
541 GTGCGGCAATCAAGAGAGCCCTGTGTAGCAACCGGCGCGAGCAACCGTGTGTAGAGAG 600
541 GTGCGGCAATCAAGAGAGCCCTGTGTAGCAACCGGCGCGAGCAACCGTGTGTAGAGAG 600
601 CTGATGCGGCGGCAACCCCGGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGC 660
601 CTGATGCGGCGGCAACCCCGGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGC 660
661 ACCCTGAATCTTCCCATCAGCCCATCAGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
661 ACCCTGAATCTTCCCATCAGCCCATCAGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
721 GACGCGCCCAAGGTGAAGAGTGGCCCTGAACGAGAGAGATCAAGGCGCTTGAACCGGC 780
721 GACGCGCCCAAGGTGAAGAGTGGCCCTGAACGAGAGAGATCAAGGCGCTTGAACCGGC 780
781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGGAGAACCTCTAC 840
781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATGCGGCCCGGAGAACCTCTAC 840
841 AACACCCCGGTGTGCGCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGGTGAAC 900
841 AACACCCCGGTGTGCGCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGGTGAAC 900
901 TTCCGCGAGCTGAACAGCGGCAACCAAGACTTCTGTGAGAGTGAAGCTGGGCAATCCCCAC 960
901 TTCCGCGAGCTGAACAGCGGCAACCAAGACTTCTGTGAGAGTGAAGCTGGGCAATCCCCAC 960
961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGAACGTTGAGCGGCAACCGCTTCACTTC 1020
961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGAACGTTGAGCGGCAACCGCTTCACTTC 1020
1021 AGCGTGCCTGTGAACAGAGACTTCCGCAAGTACACCGCTTCACTTCAGCATCCCGAGATCAAC 1080
1021 AGCGTGCCTGTGAACAGAGACTTCCGCAAGTACACCGCTTCACTTCAGCATCCCGAGATCAAC 1080
1081 AACGAGACCCCGGCAATCGGCTACAGTAAAGTGTCTGCCCCAGGCGTGAAGGCGAGC 1140
1081 AACGAGACCCCGGCAATCGGCTACAGTAAAGTGTCTGCCCCAGGCGTGAAGGCGAGC 1140
1141 CCCAGCATCTTCAAGAGAGAGATCAAGATCTGTGAACCTTCCGCGCGCGCAACCCC 1200
1141 CCCAGCATCTTCAAGAGAGAGATCAAGATCTGTGAACCTTCCGCGCGCGCAACCCC 1200
1201 GAGATGTATCTTCAAG 1260
1201 GAGATGTATCTTCAAG 1260
1261 GCGCGCAAGATCAAG 1320
1261 GCGCGCAAGATCAAG 1320

1261 GCGCGCAAGATCAAG 1320
1321 AAG 1380
1321 AAG 1380
1381 TGAACGAG 1440
1381 TGAACGAG 1440
1441 AAGCTGTGTGAG 1500
1441 AAGCTGTGTGAG 1500
1501 CTGTGAAG 1560
1501 CTGTGAAG 1560
1561 GCGGAGCTGAG 1620
1561 GCGGAGCTGAG 1620
1621 TACGAG 1680
1621 TACGAG 1680
1681 TACGAG 1740
1681 TACGAG 1740
1741 ACCGCGCAACCAAG 1800
1741 ACCGCGCAACCAAG 1800
1801 AGCATGTGATCTGAGGAG 1860
1801 AGCATGTGATCTGAGGAG 1860
1861 GAGAGCTGTGTGAG 1920
1861 GAGAGCTGTGTGAG 1920
1921 ACCCGCGGCTGTGAG 1980
1921 ACCCGCGGCTGTGAG 1980
1981 ACCCTTCAAGTGAAG 2040
1981 ACCCTTCAAGTGAAG 2040
2041 ACCGAGCGGCGGAG 2100
2041 ACCGAGCGGCGGAG 2100
2101 CTGAG 2160
2101 CTGAG 2160
2161 AGCGAG 2220
2161 AGCGAG 2220
2221 AACGAGATCAATGAG 2280
2221 AACGAGATCAATGAG 2280
2281 CAG 2340
2281 CAG 2340
2341 GTGCTGTCTGTGAG 2400
2341 GTGCTGTCTGTGAG 2400

QY 2401 TACGTGGGAGGGGGCCCTTAGATCGATTAAAGCTTCCGGGGCTAGACCCGGTGA 2460
DB 2401 TACGTGGGAGGGGGCCCTTAGATCGATTAAAGCTTCCGGGGCTAGACCCGGTGA 2460
QY 2461 TTC 2463
DB 2461 TTC 2463
RESULT 2
AX455914 2469 bp DNA linear PAT 06-JUL-2002
LOCUS Sequence 30 from Patent WO0204493.
DEFINITION AX455914
ACCESSION AX455914 GI:21714899
KEYWORDS
SOURCE synthetic construct
ORGANISM other sequences; artificial sequences.
REFERENCE 1
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Patent: WO 0204493-A 30 17-JAN-2002;
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source 1. 2469
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975(+)"
ORIGIN
Query Match 99.2%; Score 2442.2; DB 6; Length 2469;
Best Local Similarity 99.6%; Pred. No. 1.2e-200;
Matches 2460; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGAGCCGACCAATGAGCCGACATGAGCCGACGACCAAGCCCAATCTCTGATGAG 60
DB 1 GTGAGCCGACCAATGAGCCGACATGAGCCGACGACCAAGCCCAATCTCTGATGAG 60
QY 61 CGGAGCACTTCAAGGGCCCAAGCCGATCATCAATGCTTCACTGCGGCAAGAGGGC 120
DB 61 CGGAGCACTTCAAGGGCCCAAGCCGATCATCAATGCTTCACTGCGGCAAGAGGGC 120
QY 121 CACATGCGCGCACTGCGCGCGCGCGCGCAAGAGGGCTGAGAGTGGCGGCAAGAG 180
DB 121 CACATGCGCGCACTGCGCGCGCGCGCGCAAGAGGGCTGAGAGTGGCGGCAAGAG 180
QY 181 GGGCAACAGATGAGAGCTGACCGAGCGCCAGGCCCAACTTCTTCCGAGAGACTGGCC 240
DB 181 GGGCAACAGATGAGAGCTGACCGAGCGCCAGGCCCAACTTCTTCCGAGAGACTGGCC 240
QY 241 TTCTCCCAAGGAGCGCGCGAGTTTCCCAAGGAGAGACCGCGCCCAACAGCCCAACC 300
DB 241 TTCTCCCAAGGAGCGCGCGAGTTTCCCAAGGAGAGACCGCGCCCAACAGCCCAACC 300
QY 301 AGCGGAGAGCTGAGGTGGCGCGCGCAACCCCGGAGAGGGCGCGCGCGCGCGCGAG 360
DB 301 AGCGGAGAGCTGAGGTGGCGCGCGCAACCCCGGAGAGGGCGCGCGCGCGCGCGAG 360
QY 361 GGCACCTGAACTTCCCGAGATCACTCTGTGAGAGCGCCCTTGTGAGAGCATCAAGTG 420
DB 361 GGCACCTGAACTTCCCGAGATCACTCTGTGAGAGCGCCCTTGTGAGAGCATCAAGTG 420
QY 421 GGGGCGCAAGATCAAGAGGGCTGTGTGAGACCGGGCGCGAGCAACCGTGTGAGAGAG 480
DB 421 GGGGCGCAAGATCAAGAGGGCTGTGTGAGACCGGGCGCGAGCAACCGTGTGAGAGAG 480
QY 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGGGGCGGCAATGGCGGCTTCAATAG 540
DB 481 ATGAGGCTGCGCGGCAAGTGAAGCCCAAGATGATGGGGCGGCAATGGCGGCTTCAATAG 540

QY 541 GTGGCGCAGTACGACCAAGATCTGATCGAGCGCAAGAGCCATCGGCAACCGTG 600
DB 541 GTGGCGCAGTACGACCAAGATCTGATCGAGCGCAAGAGCCATCGGCAACCGTG 600
QY 601 CTGATCGGCGCCCAAGCGCGTGAACATCATCGGCGGCAACATCTGACCGAGCTGGCTTC 660
DB 601 CTGATCGGCGCCCAAGCGCGTGAACATCATCGGCGGCAACATCTGACCGAGCTGGCTTC 660
QY 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATG 720
DB 661 ACCCTGAATCTTCCCATGAGCCCATCGAGACCGTGGCCGTGAAGCTGAAGCCCGCATG 720
QY 721 GACGCGCCCAAGATGAGAGTGGCGCGCGTGAACCGGAGAGAAATGAAAGGCGCTTGAACCG 780
DB 721 GACGCGCCCAAGATGAGAGTGGCGCGCGTGAACCGGAGAGAAATGAAAGGCGCTTGAACCG 780
QY 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACCAAGATGAGCGCCCGAGAACCCCTAC 840
DB 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACCAAGATGAGCGCCCGAGAACCCCTAC 840
QY 841 AACACCCCGTGTTCGCTCATCAAGAGAGAGAGCAAGCAAGTGGCGCAAGCTGTGAG 900
DB 841 AACACCCCGTGTTCGCTCATCAAGAGAGAGAGCAAGCAAGTGGCGCAAGCTGTGAG 900
QY 901 TTCCGAGCTGAACAAGCGCACCCAGACTTCTGAGAGGTGAGCTGGGCAATCCCCAC 960
DB 901 TTCCGAGCTGAACAAGCGCACCCAGACTTCTGAGAGGTGAGCTGGGCAATCCCCAC 960
QY 961 CCGCGCGCTGAAGAGAGAGAGCGGAGCGGTGAGAGTGGGGGAGCGCTTCTTC 1020
DB 961 CCGCGCGCTGAAGAGAGAGAGCGGAGCGGTGAGAGTGGGGGAGCGCTTCTTC 1020
QY 1021 AGCGTCCCTTGAAGAGAGCTTCCGCAAGTACACCGCTTCAACATCCCAAGCATCAAC 1080
DB 1021 AGCGTCCCTTGAAGAGAGCTTCCGCAAGTACACCGCTTCAACATCCCAAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCAATCCGCTTCAAGTACAAAGTCTGCGCGCGGCTGAGAGGGCAGC 1140
DB 1081 AACGAGACCCCGGCAATCCGCTTCAAGTACAAAGTCTGCGCGCGGCTGAGAGGGCAGC 1140
QY 1141 CCGAGATCTTCAAGAGAGAGTCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCC 1200
DB 1141 CCGAGATCTTCAAGAGAGAGTCAAGATCTGAGAGCGCTTCCGCGCGCGCAACCCC 1200
QY 1201 GAGATGATCTACCA-----GGCGCCCTGTAGTGGGCGAGCACTGAGATGAGC 1254
DB 1201 GAGATGATCTACCAAGATCTGAGAGAGCTGTAGCTGTAGCTGGGCGAGCATCTGGC 1260
QY 1255 CAGCACCGGCGCAAGATGAGAGAGCTGTGCAAGACCTGTGCGCTGGGGCTTCAACACC 1314
DB 1261 CAGCACCGGCGCAAGATGAGAGAGCTGTGCAAGACCTGTGCGCTGGGGCTTCAACACC 1320
QY 1315 CCGGACAAAGAGCAACGAGAGAGCGCCCTTCTGTGATGAGGTGAGTGAAGCTGCAACCC 1374
DB 1321 CCGGACAAAGAGCAACGAGAGAGCGCCCTTCTGTGATGAGGTGAGTGAAGCTGCAACCC 1380
QY 1375 GACAAATGAGACGCTGAGCGCATCGAGCTGCCGAGAGAGAGAGTGAACCGTGAAGCAGC 1434
DB 1381 GACAAATGAGACGCTGAGCGCATCGAGCTGCCGAGAGAGAGAGTGAACCGTGAAGCAGC 1440
QY 1435 ATTCAGAACTGTGTGGCAAGCTGAATCTGGGCGCAGCTTAACCCCGCATCAAGTG 1494
DB 1441 ATTCAGAACTGTGTGGCAAGCTGAATCTGGGCGCAGCTTAACCCCGCATCAAGTG 1500
QY 1495 CCGCAGCTGTGAAGTGTGTGGCGCGCGCGCAAGCGCATCGAGCAATCGTGGCCCGTGAAC 1554
DB 1501 CCGCAGCTGTGAAGTGTGTGGCGCGCGCGCAAGCGCATCGAGCAATCGTGGCCCGTGAAC 1560
QY 1555 GAGAGGCGGAGCTGAGCTGGCGAGAACCGGAGATCTGTGCGGAGCGCGCTGCAACGCG 1614
DB 1561 GAGAGGCGGAGCTGAGCTGGCGGAGAACCGGAGATCTGTGCGGAGCGCGCTGCAACGCG 1620

QY 1615 GTGATCTAGACCCGAGCAAGGACCTGTGTGCGGAGATCCAGAACGAGGCGACGACG 1674
Db 1621 GTGTACTAGACCCGAGCAAGGACCTGTGTGCGGAGATCCAGAACGAGGCGACGACG 1680
QY 1675 TGGACCTTACGAGATCTTACGAGAGGCTTCAAGAACCTGAGAGACCGGCAAGTACCGCAG 1734
Db 1681 TGGACCTTACGAGATCTTACGAGAGGCTTCAAGAACCTGAGAGACCGGCAAGTACCGCAG 1740
QY 1735 ATGGCAGCCGCGCAACGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1794
Db 1741 ATGGCAGCCGCGCAACGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1795 ATGAGAGAGATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAG 1854
Db 1801 ATGAGAGAGATGTGATCTGTGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAG 1860
QY 1855 ACCTGGAGAGACCTGTGAGACCGAGCTTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1914
Db 1861 ACCTGGAGAGACCTGTGAGACCGAGCTTCTGTGGAGAGAGAGAGAGAGAGAGAGAGAG 1920
QY 1915 GTGAAACCCCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
Db 1921 GTGAAACCCCGCGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
QY 1975 GCGGAGACCTTCTAGGTGAGACCGGCGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAG 2034
Db 1981 GCGGAGACCTTCTAGGTGAGACCGGCGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAG 2040
QY 2035 TAGGTGAGACCGGCGCGCGAG 2094
Db 2041 TAGGTGAGACCGGCGCGCGAG 2100
QY 2095 ACCGAGCTGAG 2154
Db 2101 ACCGAGCTGAG 2160
QY 2155 ACCGAGCTGAG 2214
Db 2161 ACCGAGCTGAG 2220
QY 2215 CTGTGAGACGAGATCATGAG 2274
Db 2221 CTGTGAGACGAGATCATGAG 2280
QY 2275 CCGGCGCAACGAG 2334
Db 2281 CCGGCGCAACGAG 2340
QY 2335 CCGAGAGTGTCTGTCTGTGAG 2394
Db 2341 CCGAGAGTGTCTGTCTGTGAG 2400
QY 2395 GACCTGTGAG 2454
Db 2401 GACCTGTGAG 2460
QY 2455 GGTGAAATTC 2463
Db 2461 GGTGAAATTC 2469

RESULT 3
AX455916 2457 bp DNA linear PAT 06-JUL-2002
LOCUS AX455916
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901
KEYWORDS
SOURCE
ORGANISM synthetic construct
other sequences; artificial sequences.
REFERENCE
AUTHORS zur Megeide, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, E.

TITLE polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Polypeptides and uses thereof
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
FEATURES
source
1. 2457
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YMMM"
ORIGIN
Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 4e-200;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;
QY 1 GTGAGCGCACCATGAGCGGAG 60
Db 1 GTGAGCGCACCATGAGCGGAG 60
QY 61 CGGAGCACTTCAAGAGGCGCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
Db 61 CGGAGCACTTCAAGAGGCGCCCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 121 CACATGCGCCGCACTGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
Db 121 CACATGCGCCGCACTGCGCGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 GCGCAGAGATGAG 240
Db 181 GCGCAGAGATGAG 240
QY 241 TTCCCGCGAG 300
Db 241 TTCCCGCGAG 300
QY 301 AGCGCGAG 360
Db 301 AGCGCGAG 360
QY 361 GCGACCTGAACTTCCCGAG 420
Db 361 GCGACCTGAACTTCCCGAG 420
QY 421 GCGCGCGAGATCAAG 480
Db 421 GCGCGCGAGATCAAG 480
QY 481 ATGAGCTGTGCGCGAG 540
Db 481 ATGAGCTGTGCGCGAG 540
QY 541 GTGCGCGAGATCAAG 600
Db 541 GTGCGCGAGATCAAG 600
QY 601 CTGATGCGCGCGAG 660
Db 601 CTGATGCGCGCGAG 660
QY 661 ACCCTGAACTTCCCGAG 720
Db 661 ACCCTGAACTTCCCGAG 720
QY 721 GACGCGCGCGAG 780
Db 721 GACGCGCGCGAG 780
QY 781 ATCTGCGAGAGAGATGAG 840
Db 781 ATCTGCGAGAGATGAG 840
QY 841 AACACCCCGGTGTGCGCATCAAG 900

Db 841 AACACCCCTGTTGGCCATCAAGAAAGACAGACCAAGTGGGCGACGCTGGAGAC 900
Qy 901 TTCCGAGACTGAACAAGCGACCCAGAGCTTCTGGAGGTGACGTGGGCAATCCCCAC 960
Db 901 TTCCGAGACTGAACAAGCGACCCAGAGCTTCTGGAGGTGACGTGGGCAATCCCCAC 960
Qy 961 CCCGCGGCTTGAAGAAAGAAAGAGAGCGTGAACGTGCTGAGAGGTGGGCAATCTTC 1020
Db 961 CCCGCGGCTTGAAGAAAGAAAGAGAGCGTGAACGTGCTGAGAGGTGGGCAATCTTC 1020
Qy 1021 AGCGTGGCTTGAAGAGAGCTTCCGCAAGTACACCGCTTCAACATCCCGACATCAAC 1080
Db 1021 AGCGTGGCTTGAAGAGAGCTTCCGCAAGTACACCGCTTCAACATCCCGACATCAAC 1080
Qy 1081 AACGAGACCCCGGCAATCCGCTAACAGTACAAAGTGTGCTGCGCCAGGGCTGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCAATCCGCTAACAGTACAAAGTGTGCTGCGCCAGGGCTGAAGGGCAGC 1140
Qy 1141 CCCAGATCTTCAAGAGCAATGACCAAGATCTTGAAGCTTCCGCGCCGCAACCCG 1200
Db 1141 CCCAGATCTTCAAGAGCAATGACCAAGATCTTGAAGCTTCCGCGCCGCAACCCG 1200
Qy 1201 GAGATGTATCTAACAGAGGCCCCCTGTATGCTGGAGAGGAGCTTGAAGATCCGACAGCAG 1260
Db 1201 GAGATGTATCTAACAGAGGCCCCCTGTATGCTGGAGAGGAGCTTGAAGATCCGACAGCAG 1260
Qy 1261 CGCGCAAGATCGAGAGGCTGCGCAAGCACTGCTGCGCTGAGGCTTCAACACCCCGAC 1320
Db 1261 CGCGCAAGATCGAGAGGCTGCGCAAGCACTGCTGCGCTGAGGCTTCAACACCCCGAC 1320
Qy 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCGCAAG 1380
Db 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCGCAAG 1380
Qy 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCGCAAG 1374
Db 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGATGGGCTACGAGCTGCAACCCGCAAG 1374
Qy 1381 TGGACCTGACAGCCATCGAGCTGCGGAGAGAGAGAGCTGAGCGGTGAACGATCCAG 1440
Db 1375 TGGACCTGACAGCCATCGAGCTGCGGAGAGAGAGAGCTGAGCGGTGAACGATCCAG 1434
Qy 1441 AAGCTGTGGGCAAGCTGAATCTGGGCAAGCAGATCTAACCCCGGCAATCAAGGTGCGCAG 1500
Db 1435 AAGCTGTGGGCAAGCTGAATCTGGGCAAGCAGATCTAACCCCGGCAATCAAGGTGCGCAG 1494
Qy 1501 CTGTGCAAGCTGCTGCGCGGCGCAAGGCTTGAACGATCTGCTGCTGACCGAGAG 1560
Db 1495 CTGTGCAAGCTGCTGCGCGGCGCAAGGCTTGAACGATCTGCTGCTGACCGAGAG 1554
Qy 1561 GCGGAGCTGAGCTGCGCGGCAAGCAGGAGATCTGCGGAGGCGGTGAACGCGGTGAC 1620
Db 1555 GCGGAGCTGAGCTGCGCGGCAAGCAGGAGATCTGCGGAGGCGGTGACGCGGTGAC 1614
Qy 1621 TACGACCCCGCAAGAGCTGTGGCGAGATCCAGAGAGGCGCGCAAGCTGAGAC 1680
Db 1615 TACGACCCCGCAAGAGCTGTGGCGAGATCCAGAGAGGCGCGCAAGCTGAGAC 1674
Qy 1681 TACGAGATCTAACAGAGGCTTCAAGAGCTTGAAGCGGCGAGTACCGCAAGATGCGC 1740
Db 1675 TACGAGATCTAACAGAGGCTTCAAGAGCTTGAAGCGGCGAGTACCGCAAGATGCGC 1734
Qy 1741 ACCGCGCAACCAAGAGCTGAGAGCTGAGCGAGGCGGTGAAGAAATCGCATGGAG 1800
Db 1735 ACCGCGCAACCAAGAGCTGAGAGCTGAGCGAGGCGGTGAAGAAATCGCATGGAG 1794
Qy 1801 AGCATGTGATCTGGGCAAGACCCCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 1795 AGCATGTGATCTGGGCAAGACCCCAAGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1854
Qy 1861 GAGACTGTGAGACCGACTTCTGAGAGGCGCACTGAGTCCCGAGTGGGAGTTCTGAGAC 1920
Db 1855 GAGACTGTGAGACCGACTTCTGAGAGGCGCACTGAGTCCCGAGTGGGAGTTCTGAGAC 1914
Qy 1921 ACCCCCCCTGTGAGAGCTGTGTACAGCTGAGAGAGGCGCATCATCGGCGCGAG 1980

Db 1915 ACCCCCCCTGTGAGAGCTGTGTACAGCTGAGAGAGAGGCGCATCATCGGCGCGAG 1974
Qy 1981 ACCCTTACGTGAGAGGCGCGCGCAACCGCGAGACCAAGATGGGCAAGGCGGCTACGTC 2040
Db 1975 ACCCTTACGTGAGAGGCGCGCGCAACCGCGAGACCAAGATGGGCAAGGCGGCTACGTC 2034
Qy 2041 ACCGACCGGCGCGCGCAAGAGTGTGAGGCTGAGCGGAGACCGAGACCAAGAGAGCGAG 2100
Db 2035 ACCGACCGGCGCGCGCAAGAGTGTGAGGCTGAGCGGAGACCAAGAGAGCGAG 2094
Qy 2101 CTGAGAGCATCAAGCTGCGCTTGAAGAGCAAGCGGCAAGGAGTGAATCTGACCGAC 2160
Db 2095 CTGAGAGCATCAAGCTGCGCTTGAAGAGCAAGCGGCAAGGAGTGAATCTGACCGAC 2154
Qy 2161 AGCGATACGCTTGGGCAATCTCAAGGCGCGCGAGCGGCAAGAGGAGGAGGCTGTGTC 2220
Db 2155 AGCGATACGCTTGGGCAATCTCAAGGCGCGCGAGCGGCAAGAGGAGGAGGCTGTGTC 2214
Qy 2221 AACGATCATGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGCGCGC 2280
Db 2215 AACGATCATGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGCGCGC 2274
Qy 2281 CAAAGGAGCATCGCGCGCAACGAGAGATCGAACAGCTGTGAGCAAGGCGCATCCGCAAG 2340
Db 2275 CAAAGGAGCATCGCGCGCAACGAGAGATCGAACAGCTGTGAGCAAGGCGCATCCGCAAG 2334
Qy 2341 GTGCTGTCTTGAAGCGGCAATCGATGCGGCGCATCTGATTTACAGTATCAAGACGACTG 2400
Db 2335 GTGCTGTCTTGAAGCGGCAATCGATGCGGCGCATCTGATTTACAGTATCAAGACGACTG 2394
Qy 2401 TACGTGGGAGGCGGCGCGCTGAGATCGATTAAGCTTCCGCGGCTGAGCAAGGAGTGA 2460
Db 2395 TACGTGGGAGGCGGCGCGCTGAGATCGATTAAGCTTCCGCGGCTGAGCAAGGAGTGA 2454
Qy 2461 TTC 2463
Db 2455 TTC 2457

RESULT 4
BD263704
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITILE
JOURNAL
COMMENT

BD263704 2306 bp DNA linear PART 17-JUL-2003
Improved expression of HIV polypeptides and production of
virus-like particles.
BD263704
BD263704.1 GI:33073472
JP 200253124-A/71.
KEYWORDS
JP 200253124-A/71.
SOURCE
synthetic construct
ORGANISM
other sequences; artificial sequences.
1 (bases 1 to 2306)
Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
Improved expression of HIV polypeptides and production of
virus-like particles
Patent: JP 200253124-A 71 08-OCT-2002;
CHIRON CORP
OS Artificial Sequence
PN JP 200253124-A/71
PD 08-OCT-2002
PR 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIWASTAVA, YING LIAN, PI
KARIN HARTOG,
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).promod.Rtopc.YM FH Key
FT source
FT /organism='Artificial Sequence'.
FT Location/Qualifiers

```
source 1..2306
/mol_type="synthetic construct"
/db_xref="taxon:32630"

Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

ORIGIN

170 GCGGCAAGAGAGGCCCAACGATGAAAGATGCAACGAGCCGCAACTTCTCCGCG 229
    1 GCGGCGCGCAAGAGACCAAAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTCCGCG 60
QY 230 AGCACTGGGCTTCCCGCAAGGAGCGCGGAGTTCCCGCAAGAGACCGCGCA 289
    61 AAGACCTGGGCTTCTGCAAGGAGAGCCCGGAGTTGACAGACGAGACCGCGCA 120
QY 290 ACAGCCCAACAGCCGCGAGCTGCAAGTGCAGCGCG-----ACAACCCCGCAGCGAG 343
    121 ACAGCCCAACCGCGCGAGCTGCAAGTGTGGGCGCGAGAACACACCTGAGCGAG 180
QY 344 CCGGCGCGAGCGCCGCAAGGCACTG-----AATTCCCGCAATCACTTGTGGCAGC 397
    181 CCGGCGCGAGCGCCGCAAGGCACTTCAACTTCCCGCAATCACTTGTGGCAGC 240
QY 398 GCGCCCTGGTGAAGCATCAAGTGGGCGGCGAGATCAAGAGGCGCTGTGCAACCGCG 457
    241 GCGCCCTGGTGAACATCAAGATGGGCGGCGAGCTCAAGAGGCGCTGTGCAACCGCG 300
QY 458 CCGAGCAACACGCTGTGAGAGAGATGAGCTGCCCGCAAGTGAAGCCCAAGATGATCG 517
    301 CCGAGCAACACGCTGTGAGAGAGATGAACTTGC CGGCAAGTGAAGCCCAAGATGATCG 360
QY 518 GCGGCAATGGCGGCTTCAATCAAGTGGCGCAAGTACGACCAAGATCTGTATGAGATCTGCG 577
    361 GCGGCAATGGCGGCTTCAATCAAGTGGCGCAAGTACGACCAAGATCTGTATGAGATCTGCG 420
QY 578 GGAAGAGGCGCATTCGGGCAACGATGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 637
    421 GCGACAGAGCGCATTCGGGCAACGATGAGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCG 480
QY 638 ACATGTGACCCAGCTGTGAGTCACTTCCCGCATCAAGCCCATCGAGACCGTGG 697
    481 ACCTGTGACCCAGATTCGGCTGCACTTCACTTCCCGCATCAAGCCCATCGAGACCGTGG 540
QY 698 CCGTGAAGCTGAAGCCCGGCAATGAGTGGCGCAAGTGAAGCATGTGGCCCTGACCGAG 757
    541 CCGTGAAGCTGAAGCCCGGCAATGAGTGGCGCAAGTGAAGCATGTGGCCCTGACCGAG 600
QY 758 AGAAGATCAAGGCGCTGACCGGCAATCTGAGAGAGATGAGAGAGGCGCAAGATCAACA 817
    601 AGAAGATCAAGGCGCTGAGTGAAGATCTGACCGAGATGAGAGAGAGGCGCAAGATCAACA 660
QY 818 AGATCGGCGCGCGAGAACCCCTTCAACACCCCGTGTTCGCAATCAAGAGAGAGACAGA 877
    661 AGATCGGCGCGCGAGAACCCCTTCAACACCCCGTGTTCGCAATCAAGAGAGAGACAGA 720
QY 878 CCAAGTGGCGAGCTGTGACTTCCGAGAGTGAACAAGCGCACCCGACCTTCTGG 937
    721 CCAAGTGGCGAGCTGTGACTTCCGAGAGTGAACAAGCGCACCCGACCTTCTGG 780
QY 938 AGGTGAGCTGGGCAATCCCGCAACCCCGCGCTGAAAGAAAGAAAGACGTGACGTGC 997
    781 AGGTGAGCTGGGCAATCCCGCAACCCCGCGCTGAAAGAAAGAAAGACGTGACGTGC 840
QY 998 TGAAGCTGGGCGAGCGCTTCAAGCGTCCCTGACAGAGACTTCGCAAGTACACCG 1057
    841 TGAAGCTGGGCGAGCGCTTCAAGCGTCCCTGACAGAGACTTCGCAAGTACACCG 900
QY 1058 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCAATCGCTACAGTACAGGTGC 1117
    901 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCAATCGCTACAGTACAGGTGC 960
```

```
QY 1118 TGCCCCAGGCGTGAAGGAGGAGCCCGCAGATCTTCCAGAGAGCATGACCAAGATCTTGG 1177
    961 TGCCCCAGGCGTGAAGGAGGAGCCCGCAGATCTTCCAGAGAGCATGACCAAGATCTTGG 1020
QY 1178 AGCCCTTCCGCGCGCGCAGACCCCGAGATCTGATCTTCAAGAGGCGCCCTGTACGTGGGCA 1237
    1021 AGCCCTTCCGAGAGAGAACCCCGCATCTGTATCTTACAGAGCGCCCTGTACGTGGGCA 1080
QY 1238 GCGACCTTGAAGATCGGCGAGACCCGCGCAGATGAGAGAGCTTGGCAAGCACCTGTGC 1297
    1081 GCGACCTTGAAGATCGGCGAGACCCGCGCAGATGAGAGAGCTTGGCGCAGCACCTGTGC 1140
QY 1298 GCTGGGCGCTTCAACGACCCCGCAGAGAGACCAAGAGAGGCGCCCTTCCGTGGAATGG 1357
    1141 GCTGGGCGCTTCAACGACCCCGCAGAGAGACCAAGAGAGGCGCCCTTCCGTGGAATGG 1200
QY 1358 GCTACGAGCTGCAACCCCGCAAGTGAACGTGCAAGCCCATCGAGCTGCGCGAGAGAGA 1417
    1201 GCTACGAGCTGCAACCCCGCAAGTGAACGTGCAAGCCCATCGAGCTGCGCGAGAGAGA 1260
QY 1418 GCTGGAACGTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACGAGATCT 1477
    1261 GCTGGAACGTGAAACGACATCCAGAAAGCTGTGGGCAAGCTGAACTGGGCGACGAGATCT 1320
QY 1478 ACCCGGAGATCAAGGTGGGCGGAGCTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGACCG 1537
    1321 ACCCGGAGATCAAGGTGAAAGAGCTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGACCG 1380
QY 1538 ACATGTGCGCCCTGACCGAGAGAGCGCGAGCTGAGCTGGCGCAAGACCGGAGATCTTGC 1597
    1381 AGGTATCCCTTGAACCGAGAGAGCGCGAGCTGAGCTGGCGGAGAACCGGAGATCTTGA 1440
QY 1598 GCGAGCCGTGCAACGCGCTGTACTAGACCCCGCAAGAGACTTGTGGCGAGATTCAGA 1657
    1441 AGGAGCCGTGCAACGAGGTGTACTAGACCCCGCAAGAGACTTGTGGCGAGATTCAGA 1500
QY 1658 AGCAGGCGCAAGACCAAGTGAACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGAAAGA 1717
    1501 AGCAGGCGCAAGGCGCAAGTGAACCTTACAGATCTTACAGAGGCGCTTCAAGAACCTGAAAGA 1560
QY 1718 CCGGCAAGTACGCAAGATGCGCACCGGCGCACCAACGACCTGAAAGAGCTGACCGAG 1777
    1561 CCGGCAAGTACGCGCATGTGGCGCGCGCGCACCAACGAGCTGAAAGAGCTGACCGAG 1620
QY 1778 CCGTGAAGATGCGCATGAGAGCATGTGTATCTTGGGCAAGACCCCGAAGTTCCGCG 1837
    1621 CCGTGAAGATGAGCATGAGAGCATGTGTATCTTGGGCAAGATCCCGAAGTTCAAGC 1680
QY 1838 TGCCATCCAGAGAGGACCTTGGAGACCTTGTGACCGACTTACCTGAGGAGGCACTTGA 1897
    1681 TGCCATCCAGAGAGGACCTTGGAGAGCTTGTGATGAGATCTTGGAGGAGGCACTTGA 1740
QY 1898 TCCCGAGTGGAGTTCGTGAACACCCCGCGCTGTGTAAGCTGTGTATCAAGCTGAGAGA 1957
    1741 TCCCGAGTGGAGTTCGTGAACACCCCGCGCTGTGTAAGCTGTGTATCAAGCTGAGAGA 1800
QY 1958 AGGAGCCATCATCGGCGCGAGACCTTCTTACGTGAGCGGCGCGCAACCGCGAGACCA 2017
    1801 AGGAGCCATCATCGTGGCGCGCGAGACCTTCTTACGTGAGCGGCGCGCAACCGCGAGACCA 1860
QY 2018 AGATGGGAGAGGCGCGCTTACGTGAACCGAGCGGCGCGCGCAAGATCGTGAACCTGACCG 2077
    1861 AGCTGGGAGAGGCGCGCTTACGTGAACCGAGCGGCGCGCGCAAGATCGTGAACCTGACCG 1920
QY 2078 AGACCAACCAACCAAGAGACCGAGCTGAGGCGCATCAAGCTGGCCCTGCGAGACAGCGGA 2137
    1921 AGACCAACCAACCAAGAGACCGAGCTGAGGCGCATCAAGCTGGCCCTGCGAGACAGCGGC 1980
QY 2138 GCGAGGTGAACATCGTGAACCGAGCGCAGTACGCTTGGGCAATCAAGGCGCAGCGCG 2197
    1981 TGAAGTGAACATCGTGAACCGAGCGCAGTACGCTTGGGCAATCAAGGCGCAGCGCG 2040
```

QY 2198 ACAAGAGGAGGAGCTGTGTAACCAATCATCGAGCACTGTATCAAGAGAGAAAG 2257
DB 2041 ACAAGAGGAGGAGCTGTGTAAGCAATCATCGAGCACTGTATCAAGAGAGAAAG 2100
QY 2258 TGTACTGTGAGCTGTGTCGCCGCCCAACAAGGATCGGCGCAACAGAGCATGCAAAAG 2317
DB 2101 TGTACTGTGAGCTGTGTCGCCGCCCAACAAGGATCGGCGCAACAGAGCATGCAAAAG 2160
QY 2318 TGTGAGCAAGGAGCATCGCAAGGTGCTGTCTGTGACGGCATCGATGCGGCGATGTA 2377
DB 2161 TGTGAGCGCCCGGAGCATCGCAAGGTGCTGTCTGTGACGGCATCGATGCGGCGATGTA 2220
QY 2378 TGTACCAAGTACATGAGACCACTGTAGTGGGCAAGGCGGCGGCTGTAGATTCATTAAGC 2437
DB 2221 TGTACCAAGTACATGAGACCACTGTAGTGGGCAAGGCGGCGGCTGTAGATTCATTAAGC 2280
QY 2438 TTCCCGGGGCTAGACCGGATGAATTC 2463
DB 2281 TTCCCGGGGCTAGACCGGATGAATTC 2306

RESULT 5
CQ870574 2306 bp DNA linear PAT 13-SEP-2004
LOCUS CQ870574
DEFINITION Sequence 82 from Patent EP1433851.
ACCESSION CQ870574
VERSION CQ870574.1 GI:52000090
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
1
AUTHORS
Barnett,S., Zuermeide,J., Srivastava,I., Lian,Y., Hartog,K.,
Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE
Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL
Patent: EP 1433851-A 82 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
1. 2306
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS (-). protomod.RTopt.YM"

ORIGIN
Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
170 GGGGCAAGAGGAGCCACGATGAAGATGACCGAGCGGCGCAACTTCTTCGGG 229
DB 1 GGGGCGCGAAGAGACCAAAATGAAGATTGCACTAGAGACAGGCTTAATTTCTTCGGG 60
QY 230 AGGACTGTGCTTTCCTCCCAAGGCGCAAGCGCGAGTTTCCCAAGCAAGAAACCGCGCA 289
DB 61 AGGACTGTGCTTTCCTCCCAAGGCGCAAGCGCGAGTTTCCCAAGCAAGAAACCGCGCA 120
QY 290 ACAAGCCCCCAAGCGCGAGCTGCAAGTGCAGCGG-----ACAACCCCCGAGAGAG 343
DB 121 ACAAGCCCCCAAGCGCGAGCTGCAAGTGCAGCGGCGCGCAAGAAACAAGCTTGAAGAG 180
QY 344 CGGGGCGAGAGCGCAAGGCAACCTG-----AACTTCCCGGAGATCAACCTGTGAGC 397
DB 181 CGGGGCGCGAAGCGCGAGGCAACCTGAACTTCACTTCCCGGAGATCAACCTGTGAGC 240
QY 398 GCGCCCTGTGATCATCAAGTGGCGCGCAATCAAGAGGCGCTGTGACACCGGCG 457
DB 241 GCGCCCTGTGATCATCAAGTGGCGCGCAAGTGGCGCGCAAGAGGCGCTGTGACACCGGCG 300
QY 458 CGGACGACACCTGTGATGAGATGAGCTGCGCGCAAGTGGAGCCCAAGATGATG 517

DB 301 CGGACGACACCTGTGAGAGATGAACCTGCCCGGCAAGTGAAGGCCCAAGATGATG 360
QY 518 GGGGATCGGGGGCTTTCATCAAGGTGGCCAGTACGACCAAGATCTTATCGAGATCTGCG 577
DB 361 GGGGATCGGGGGCTTTCATCAAGGTGGCCAGTACGACCAAGATCTTCCGGTGAATCTGCG 420
QY 578 GCAAGAGGCGATCGGCAACCGGCTGATCGGCGGCAACCGGCTGAACATCATCGGCGCA 637
DB 421 GCGCAAGGCGATCGGCAACCGGCTGATCGGCGGCAACCGGCTGAACATCATCGGCGCA 480
QY 638 ACATCTGACCAAGCTGAGGCTGACCTGAACTTCCCATGAGCCCATCGAGAGCTGTC 697
DB 481 ACCTCTGACCAAGATCGGCTGACCTGAACTTCCCATGAGCCCATCGAGAGCTGTC 540
QY 698 CGGTGAAGCTGAAGCCCGCATGACGCGCCCGCAAGTGAAGAGTGGCCCTGACCGAG 757
DB 541 CGGTGAAGCTGAAGCCCGCATGACGCGCCCGCAAGTGAAGAGTGGCCCTGACCGAG 600
QY 758 AGAAGATCAAGGCGCTGATCGGCAATCTGAGAGAGATGAGAGAGGCGCAAGATCAACA 817
DB 601 AGAAGATCAAGGCGCTGATCGGCAATCTGAGAGATGAGAGAGGCGCAAGATCAACA 660
QY 818 AGATCGGCGCGGAGAAACCTTACCAACACCCCGTGTGCGCATCAAGAGAGACAGCA 877
DB 661 AGATCGGCGCGGAGAAACCTTACCAACACCCCGTGTGCGCATCAAGAGAGACAGCA 720
QY 878 CCAAGTGGCGCAAGTGTGATCTTCCGAGCTGAACAAGGCAACCGAGCTTCTGCG 937
DB 721 CCAAGTGGCGCAAGTGTGATCTTCCGAGCTGAACAAGGCAACCGAGCTTCTGCG 780
QY 938 AGTGCAGCTGGGCACTCCCGCACCCCGCGGCTGAAGAGAGAGAGCGTGAACGTCG 997
DB 781 AGTGCAGCTGGGCACTCCCGCACCCCGCGGCTGAAGAGAGAGAGCGTGAACGTCG 840
QY 998 TGGAGTGGGCGAGCGCTACTTACAGTGCGCCCTGAGACGAGGACTTCCGCAAGTACAG 1057
DB 841 TGGAGTGGGCGAGCGCTACTTACAGTGCGCCCTGAGACGAGGACTTCCGCAAGTACAG 900
QY 1058 CTTTACCATCCCGAGCATCAACAAAGAGACCCCGCATCCGCTACAGTACAAAGTGC 1117
DB 901 CTTTACCATCCCGAGCATCAACAAAGAGACCCCGCATCCGCTACAGTACAAAGTGC 960
QY 1118 TCCCGCAAGGCTGGAAGGCGAGCCCGAGATTTTCAAGAGCATGACCAAGATCTGCG 1177
DB 961 TCCCGCAAGGCTGGAAGGCGAGCCCGAGATTTTCAAGAGCATGACCAAGATCTGCG 1020
QY 1178 AGCCCTTCGCGCGCGCAACCCCGAGATGATGATCAACGAGGCGCCCTGTAGTGGGCA 1237
DB 1021 AGCCCTTCGCGCGAGCAACCCCGAGATGATGATCAACGAGGCGCCCTGTAGTGGGCA 1080
QY 1238 GCGACTGTGATGCGGCGAGCAACCGCGCCAGATCGAGAGCTGCGCAAGCACTGCTGC 1297
DB 1081 GCGACTGTGATGCGGCGAGCAACCGCGCCAGATCGAGAGAGTGGCGCGAGCACTGCTGC 1140
QY 1298 GCTGGGGCTTCAACACCCCGCAAGAGCAACGAGAGGCGCCCTTCTGTGATGAG 1357
DB 1141 GCTGGGGCTTCAACACCCCGCAAGAGCAACGAGAGGCGCCCTTCTGTGATGAG 1200
QY 1358 GCTAGAGCTGACCCCGCAAGTGAACGTCAGCGCATTCGAGCTGCCCGAGAGAGAGA 1417
DB 1201 GCTAGAGCTGACCCCGCAAGTGAACGTCAGCGCATTCGAGCTGCCCGAGAGAGAGA 1260
QY 1418 GCTGACCGTGAACGATCCAGAACTGTGTGGCAAGCTGAAGTGGCGAGCGAGATCT 1477
DB 1261 GCTGACCGTGAACGATCCAGAACTGTGTGGCAAGCTGAAGTGGCGAGCGAGATCT 1320
QY 1478 ACCCGGATCAAGGTGGCGCACTGTGCAAGCTGTGCGCGCGCAAGGCGCTGACCG 1537
DB 1321 ACCCGGATCAAGGTGGCGCAAGCTGTGCAAGCTGTGCGCGCGCAAGGCGCTGACCG 1380
QY 1538 ACATGTGCGCGTGAACCGAGAGGCGGAGCTGAGAGCTGCGCGAGAAACCGAGATCTGCG 1597
DB 1381 ACATGTGCGCGTGAACCGAGAGGCGGAGCTGAGAGCTGCGCGAGAAACCGAGATCTGCG 1440


```

QY 1598 GCGAGCCCGTGCACGCGCGTGTACTACGACCCGAGCAAGGACTGTGGCCGAGATCCAGA 1657
DB 1441 AGGAGCCCGTGCACGAGGTGTACTACGACCCGAGCAAGGACTGTGGCCGAGATCCAGA 1500
QY 1658 AGCAGGCGCCAGCAGCAGTGTGACCTTACGAGATCTTACGAGAGCCCTTCAAGAACTTGAAGA 1717
DB 1501 AGCAGGCGCCAGGCGCGTGTGACCTTACGAGATCTTACGAGAGCCCTTCAAGAACTTGAAGA 1560
QY 1718 CCGGCAAGTACGCGCAAGAGTGGCAGCGGCCAGCAAGCAAGCAAGTGAAGGAGCTGACCGAGG 1777
DB 1551 CCGGCAAGTACGCGCGCATGCGCGCGCGGCCAGCAAGCAAGCTGAGAGCTGACCGAGG 1620
QY 1778 CCGTGCAGAGAAATCCGCAATGAGAGCATGTGATCTGGGCGCAAGAACCCCAAGTTCCCGC 1837
DB 1621 CCGTGCAGAGAGTGTAGACACCGAGAGCATGTGATCTGGGCGCAAGAAATCCCAAGTTCAAGC 1680
QY 1838 TGCCCATCCAGAGAGAGACTGTGGAGACTGTGTGACCGACTTACTGTGCGCAGGCTGAG 1897
DB 1681 TGCCCATCCAGAGAGAGACTGTGGAGCGCTGTGTGATGTGATCTGTGCGCAGGCTGAG 1740
QY 1898 TGCCCGAGTGGAGTTCGTGAAACACCCCGCGTGTGAGCTGTGTGATCTGTGCGCAGG 1957
DB 1741 TGCCCGAGTGGAGTTCGTGAAACACCCCGCGTGTGAGCTGTGTGATCTGTGCGCAGG 1800
QY 1958 AGGAGCCCATCATCGGCGCGCGAGACCTTCTACTGTGACCGCGCGCCAGCAAGCGAGACCA 2017
DB 1801 AGGAGCCCATCATCGTGGCGCGCGAGACCTTCTACTGTGACCGCGCGCCAGCAAGCGAGACCA 1860
QY 2018 AGATGCGCAGAGCGCGCTTACTGTGACCGAGCGGCGCGCGAGAAAGTCTGTGAGCTTGA 2077
DB 1861 AGCTGCGCAGAGCGCGCTTACTGTGACCGAGCGGCGCGCGAGAAAGTCTGTGAGCTTGA 1920
QY 2078 AGACCAACCAACCAAGAGACCGAGCTGCGAGGCGCATCAAGTGTGCGCTGTGAGCAGCGGCA 2137
DB 1921 AGACCAACCAACCAAGAGACCGAGCTGCGAGGCGCATCAAGTGTGCGCTGTGAGCAGCGGCA 1980
QY 2138 GCGAGGTGAACTCTGTGACCGAGCAGCAGTACGCGCTGTGCGCATCATCCAGGCCAGCCG 2197
DB 1981 TGGAAGTGTAACTCTGTGACCGAGCAGCAGTACGCGCTGTGCGCATCATCCAGGCCAGCCG 2040
QY 2198 ACAAGAGCGAGAGCGAGCTGTGTGAACGAGATCATGTGAGAGCTGTATCAAGAAAGAGAGG 2257
DB 2041 ACAAGAGCGAGAGCGAGCTGTGTGAACGAGATCATGTGAGAGCTGTATCAAGAAAGAGAGG 2100
QY 2258 TGTAAGTGTGAGTGGCGCGCGCGCGCGCAAGGCGATGGCGCGCAAGCAGAGATCGCAAGC 2317
DB 2101 TGTAAGTGTGAGTGGCGCGCGCGCGCGCGCAAGGCGATGGCGCGCAAGCAGAGATCGCAAGC 2160
QY 2318 TGGTGAGCAAGGCGCATCCGCAAGTGTGTCTGTGACCGGCGCATTCGATGCGGCGCATCGTGA 2377
DB 2161 TGGTGAGCGCGCGCGCATCCGCAAGTGTGTCTGTGACCGGCGCATTCGATGCGGCGCATCGTGA 2220
QY 2378 TCTACAGTGTACGTGACGACTGTGTACGTGTGGCGAGCGCGCGCTTGAAGATGATTAAGC 2437
DB 2221 TCTACAGTGTACGTGACGACTGTGTACGTGTGGCGAGCGCGCGCTTGAAGATGATTAAGC 2280
QY 2438 TTCCCGGCGGTGTAGCAGCGGTGTGAATTC 2463
DB 2281 TTCCCGGCGGTGTAGCAGCGGTGTGAATTC 2306

```

```

RESULT 6
LOCUS AR373387 2306 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 82 from patent US 6602705.
ACCESSION AR373387
VERSION AR373387.1 GI:40075490
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 2306)

```

```

AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like
JOURNAL particles
Patent: US 6602705-A 82 05-Aug-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source
1..2306
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 83.1%; Score 2046; DB 6; Length 2306;
Best Local Similarity 93.6%; Pred. No. 1e-166;
Matches 2159; Conservative 0; Mismatches 135; Indels 12; Gaps 2;
QY 170 GCGGCAAGAGGCGCCAGCAAGTGAAGACTGTGACCGGAGCGCAAGCTTCTTCCGG 229
DB 1 GCGGCGCGCAAGAGCAACCAATGAATGTGACTGTGAGAGAGAGCTTAATTTCTTCCGG 60
QY 230 AGGACTGTGAGCTTCCCGCAGGCGCAAGGCCCGGAGTTCCCGCAGAGCAAGAACCGCGCA 289
DB 61 AGGACTGTGAGCTTCCCGCAGGCGCAAGGCCCGGAGTTCAAGAGCGAGAGACCGCGCA 120
QY 290 ACAAGCCCAACAGCCGCAAGCTGTGAGGTGCGGCGG-----ACAAGCCCGCAGCGAG 343
DB 121 ACAAGCCCAACAGCCGCGAGCTGTGAGGTGCGGCGGCGGAGAAACAAGCTGAGCGAG 180
QY 344 CCGGCGCGAGCGCGAGGCGGCGGCGG-----AATTCGCCCAAGTCAAGCTGTGCGAGC 397
DB 181 CCGGCGCGAGCGCGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240
QY 398 GCCCGCTGTGTGAGCATCAAGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 457
DB 241 GCCCGCTGTGTGAGCATCAAGGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300
QY 458 CCGAGCAACCGTGTGTGAGAGATGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 517
DB 301 CCGAGCAACCGTGTGTGAGAGATGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360
QY 518 GCGGCGTGTGCGGCTTCAAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 577
DB 361 GCGGCGTGTGCGGCTTCAAGAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420
QY 578 GCAAGAGGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 637
DB 421 GCAAGAGGCGCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480
QY 638 ACATGCTGACCGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 697
DB 481 ACATGCTGACCGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540
QY 698 CCGTGAAGCTGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 757
DB 541 CCGTGAAGCTGAAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600
QY 758 AGAAGATCAAGGCGCTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 817
DB 601 AGAAGATCAAGGCGCTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 818 AGATGCGCGCGCGAGAACCTTACCAACCGCGGTGTGCGCATCAAGAAAGAGCAAGCA 877
DB 661 AGATGCGCGCGCGAGAACCTTACCAACCGCGGTGTGCGCATCAAGAAAGAGCAAGCA 720
QY 878 CCAAGTGTGCGCAAGCTGTGTGAGCTTCCCGAGCTGAACAAGGCGACCCGAGCTTGTGG 937
DB 721 CCAAGTGTGCGCAAGCTGTGTGAGCTTCCCGAGCTGAACAAGGCGACCCGAGCTTGTGG 780
QY 938 AGGTGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 997
DB 781 AGGTGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 998 TGAGCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1057

```

Db 841 TGGACGTGGGCGACGCTACTTTCAGCGTGCCTTGGACAAGACCTTCCGCACTACACCG 900
QY 1058 CCTTCAACATCCCGACATCAACACAGAACCCCGGCACTCGCTACAGTACCAAGTGC 1117
Db 901 CCTTCAACATCCCGACATCAACACAGAACCCCGGCACTCGCTACAGTACCAAGTGC 960
QY 1118 TGGCCCAAGGGCTGGAAAGGGGACGCCCAAGCATCTTCCAGAGACATGACCAAGATCTCTG 1177
Db 961 TGGCCCAAGGGCTGGAAAGGGGACGCCCAAGCATCTTCCAGAGACATGACCAAGATCTCTG 1020
QY 1178 AGCCCTTCGCGCGCCCGCAACCCCGAAGTGTGATCTTACAGAGCGCCCTGTGATGTGAGCA 1237
Db 1021 AGCCCTTCGCGCAAGACAGAACCCCGCACTGTGATCTTACAGAGCGCCCTGTGATGTGAGCA 1080
QY 1238 GCGACCTGAGATCGGCGACAGACCGGCGCAAGATCGAGAGCTGCGCAAGCATCTGCTG 1297
Db 1081 GCGACCTGAGATCGGCGACAGACCGGCGCAAGATCGAGAGCTGCGCGACAGCATCTGCTG 1140
QY 1298 GCTGGGGCTTTCACACACCCCGCAAGAGACACAGAAAGAGCGCCCTTCTGTGATG 1357
Db 1141 GCTGGGGCTTTCACACACCCCGCAAGAGACACAGAAAGAGCGCCCTTCTGTGATG 1200
QY 1358 GCTACAGAGTGCACCCCGCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1417
Db 1201 GCTACAGAGTGCACCCCGCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1260
QY 1418 GCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACAGATCT 1477
Db 1261 GCTGACCGTGAACGACATCCAGAGCTGTGGGCAAGCTGAACCTGGGCGACAGATCT 1320
QY 1478 ACCCGCGCATCAAGGTGGCGCAAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTTGAACG 1537
Db 1321 ACCCGCGCATCAAGGTGGCGCAAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTTGAACG 1380
QY 1538 ACATCGTGCCTTCAGACGAGAGGCGAGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1597
Db 1381 ACATCGTGCCTTCAGACGAGAGGCGAGCTGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGC 1440
QY 1598 GCGAGCGCGTGCACGCGGCTGTACTACAGACCCCGCAAGAGACCTGTGGCGAGATCCAGA 1657
Db 1441 AGAGCGCGCGTGCACGAGGTGTACTACAGACCCCGCAAGAGACCTGTGGCGAGATCCAGA 1500
QY 1658 AGAGGAGCGCAAGCATGTGAGCTTACAGATCTTACAGAGCGCCCTTCAAGAACCTTGAAGA 1717
Db 1501 AGAGGAGCGCGCAAGCATGTGAGCTTACAGATCTTACAGAGCGCCCTTCAAGAACCTTGAAGA 1560
QY 1718 CCGGCAAGTACCGCAAGAGGCGACCGGCGCAACCGACGAGAGCAAGCTGACCGAGG 1777
Db 1561 CCGGCAAGTACCGGCGCGATGCGGCGCGCGCAACCGACGAGAGCAAGCTGACCGAGG 1620
QY 1778 CCGTGCAGAAATCGCATGAGAGACATCGTGTGAGGAGCAAGACCCCGAAGTTCGCGC 1837
Db 1621 CCGTGCAGAAATCGGAGAGACCGGAGAGACATCGTGTGAGGAGCAAGTTCGAGTCAAGC 1680
QY 1838 TGGCCATTCAGAAAGAGACCTTGGAGAGCTGTGAGACCGACTGAGCAGGCGCACTTGA 1897
Db 1681 TGGCCATTCAGAAAGAGACCTTGGAGAGCTGTGAGAGTACTGAGCAGGCGCACTTGA 1740
QY 1898 TCCCCAGTGGAGTTCGTGAACACCCCGCGGTGAGAGCTGTGTGTAACAAGTGGAGA 1957
Db 1741 TCCCCAGTGGAGTTCGTGAACACCCCGCGGTGAGAGCTGTGTGTAACAAGTGGAGA 1800
QY 1958 AGAGCCCATCATCGGCGCGAGACCTTCTACGTGACGCGCGCGCGCAACCGCGAGACCA 2017
Db 1801 AGAGCCCATCATCGTGGGCGCGAGACCTTCTACGTGACGCGCGCGCGCAACCGCGAGACCA 1860
QY 2018 AGATCGGCAAGGCGGCTACGTGACCGACCGGCGCGCGCGCAAGAGTGTGAGCTTGAACG 2077
Db 1861 AGTGGGCGCAAGGCGGCTACGTGACCGACCGGCGCGCGCGCAAGAGTGTGAGCTTGAACG 1920
QY 2078 AGACCAACCAACGAGAGACCGAGCTGAGAGCAATCCAGCTGAGCGCTGACGAGACGAGGCGCA 2137
Db 1921 AGACCAACCAACGAGAGACCGAGCTGAGAGCAATCCAGCTGAGCGCTGACGAGACGAGGCGC 1980

QY 2138 GCGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTTGGGCAATCATCCAGGCCCAAGCCG 2197
Db 1981 TGGAGGTGAACATCTGTGACCGACAGCCAGTACGCCCTTGGGCAATCATCCAGGCCCAAGCCG 2040
QY 2198 ACAAGAGCGAGCGGAGCTGTGGAACAGATCATCGAGAGCTGATCAAGAAAGAGAGG 2257
Db 2041 ACAAGAGCGAGCGGAGCTGTGGAACAGATCATCGAGAGCTGATCAAGAAAGAGAGG 2100
QY 2258 TGTACTGAGTGTGGTCCCGCGCAAGAGGCAATCGGCGCGCAACGAGATCGACAACG 2317
Db 2101 TGTACTGAGTGTGGTCCCGCGCAAGAGGCAATCGGCGCGCAACGAGATCGACAACG 2160
QY 2318 TGTGAGCAAGGCAATCCGCAAGGTGCTTCTTGTGACGCGCATCGATGCGCGCATGTGA 2377
Db 2161 TGTGAGCGCCCGGCAATCCGCAAGGTGCTTCTTGTGACGCGCATCGATGCGCGCATGTGA 2220
QY 2378 TCTACAGTACATGAGCGACCTGTACGTGGGCGAGCGCGCGCGCTAGATCGATTAAAGC 2437
Db 2221 TCTACAGTACATGAGCGACCTGTACGTGGGCGAGCGCGCGCGCTAGATCGATTAAAGC 2280
QY 2438 TTCCCGGGGCTAGCACCGGTGAATTC 2463
Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 7
AX427930 9166 bp DNA linear PAT 20-JUN-2002
LOCUS AX427930
DEFINITION Sequence 168 from Patent W00232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.

REFERENCE
1 Huang, Y. and Nabel, G.J.
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
Patent: WO 0232943-A 168 25-APR-2002;
JOURNAL GOVERNMENT OF THE UNITED STATES (US)
Location/Qualifiers
FEATURES
source
1. .9166
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="plasmid pVR1012x/s containing HIV genes"

ORIGIN
Query Match 82.3%; Score 2027.6; DB 6; Length 9166;
Best Local Similarity 91.7%; Pred. No. 2.4e-165;
Matches 2194; Conservative 0; Mismatches 179; Indels 19; Gaps 4;

QY 14 TGGCCAGGCGCATGAGCGCAAGCGCCCAACATCTGTATGACGCGGCAACTTCA 73
Db 2961 TGGCCAGGCGCATGAGCGCAAGCGCCCAACATCTGTATGAGCGCGGCAACTTCA 3020
QY 74 AGGCGCCCAAGGCGCATCAAGTGTCTCAACTGTGCGGCAAGAGGCGCAATGCGCCGCA 133
Db 3021 AGGCGCCCAAGGCGCATCAAGTGTCTCAACTGTGCGGCAAGAGGCGCAATGCGCCGCA 3080
QY 134 ACTGCGCGCGCCCGCGCAAGAGGCGTGTGAAAGTGCAGAGAGAGGCGCACAGATGA 193
Db 3081 ACTGCGCGCGCCCGCGCAAGAGGCGTGTGAAAGTGCAGAGAGAGGCGCACAGATGA 3140
QY 194 AGACTGACACCAAGGCGCAAGGCTTCTTCCGCGAGAGACCTGTGCTTCCCAAGGCGCA 253
Db 3141 AGACTGACACCAAGGCGCAAGGCTTCTTCCGCGAGAGACCTGTGCTTCCCAAGGCGCA 3199
QY 254 AGGCGCGCGAGTTCCTCCAGAGAGAGACCGCGCGCAACAGCCCGCAAGCGCGAGCTGC 313
Db 3200 AGGCGCGCGAGTTCCTCCAGAGAGAGACCGCGCGCAACAGCCCGCAAGAGAGAGCTTC 3259

```

OY 314 AGGTGCGCG-----CGACAACCCCGAGGCGCGCGCGCGAGCGCA--- 364
DB 3260 AGTTTGGGGAAGAGACAACAATCTCTTCAGAGAGAGAGACCATAGACAAGAACTG 3319
OY 365 ----CCCTGAATTCCTCCCGATGATCCCTGTGCAAGCGCCCTGTGTAGCATCAAGTGG 421
DB 3320 TATCTTTTGTGCTCTCTCAGATCACTTTGGCAGAGAACCCCTCTGTCAATTAAGATAG 3379
OY 422 GGGGCGAGATCAAGAGAGCGCTGTGGAACAACGGCGCGGACACACCGGTGTGGAGAGA 481
DB 3380 GGGGCGAGATCAAGAGAGCGCTGTGGAACAACGGCGCGGACACACCGGTGTGGAGAGA 3439
OY 482 TGAAGCTGCGCGGCAAGTGAAGGCCAAGATATCGGCGGCAATCGGCGCTTCAACAAG 541
DB 3440 TGAACCTGCGCGGCGCTGGAAGCCAAATGATCGGCGGCAATCGGCGCTTCAACAAG 3499
OY 542 TGGCGCAGTACCAACGATCTGTATGAGATCTGCGGCAAGAGGCCATCGGACCGTGC 601
DB 3500 TGGGCGAGTACGACGATCTGTATGAGATCTGCGGCAACAGGCCATCGGACCGTGC 3559
OY 602 TGAATGCGCGCGCAACCCCGTGAATCAATGCGCGGCAACATGCTGAACCGCTGGCTGA 661
DB 3560 TGTGTGCGCGCAACCCCGTGAATCAATGCGCGGCAACATGCTGAACCGCTGGCTGA 3619
OY 662 CCTGAATCTTCCCATCAAGCCCATCGAGACGCTGCGCGTGAAGCTGAAGCCGCGCATGG 721
DB 3620 CCTGAATCTTCCCATCAAGCCCATCGAGACGCTGCGCGTGAAGCTGAAGCCGCGCATGG 3679
OY 722 AGGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGAGAAAGATCAAGCCCTGAACCGCA 781
DB 3680 AGGCGCCCAAGTGAAGCAGTGGCCCTGACCGAGAGAAAGATCAAGCCCTGTGTGAGA 3739
OY 782 TGTGAGAGAGATGAGAGAGAGAGAGATCAACCAAGATCGGCGCGGAAACCTCTACA 841
DB 3740 TGTGAGAGAGATGAGAGAGAGAGAGATCAACCAAGATCGGCGCGGAAACCTCTACA 3799
OY 842 ACAACCCCGTGTTCGCGCATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 3800 ACAACCCCGTGTTCGCGCATCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3859
OY 902 TCCGCGAGCTGAACAAGAGCAGCCAGAGACTTCTGGAGGTGACAGCTGGGCATCCCCACC 961
DB 3860 TCCGCGAGCTGAACAAGAGCAGCCAGAGACTTCTGGAGGTGACAGCTGGGCATCCCCACC 3919
OY 962 CGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1021
DB 3920 CGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3979
OY 1022 GCGTGCCTCTGAGCAGAGACTTCCGCAAGTACACGCTTCAACATCCCGACATCAACA 1081
DB 3980 GCGTGCCTCTGAGCAGAGACTTCCGCAAGTACACGCTTCAACATCCCGACATCAACA 4039
OY 1082 AGGAGACCCCGGCAATCGGCTACAGTACAGTGTGCTCCAGAGGCTGAAGAGAGAGAG 1141
DB 4040 AGGAGACCCCGGCAATCGGCTACAGTACAGTGTGCTCCAGAGGCTGAAGAGAGAGAG 4099
OY 1142 CGAGATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
DB 4100 CGGCAATCTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4159
OY 1202 AGATGATGATCAACA-----GACCCCGCTGTACGTGGGAGAGAGAGAGAGAGAGAG 1255
DB 4160 ACATGATGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4219
OY 1256 AGACCGCGCGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1315
DB 4220 AGACCGCGCGCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4279
OY 1316 CGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1375
DB 4280 CGGACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4339

```

```

OY 1376 ACAAGTGAACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435
DB 4340 ACAAGTGAACGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4399
OY 1436 TCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAAGATCTTACCCCGCATCAAGAGTGC 1495
DB 4400 TCCAGAGCTGTGGGCAAGCTGAACTGGGCGCAAGATCTTACCCCGCATCAAGAGTGC 4459
OY 1496 GCGAGCTGTGAGAGCTGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1555
DB 4460 GCGAGCTGTGAGAGCTGTGGGCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4519
OY 1556 AGAGAGCGAGCTGAGAGCTGGCGGAGAAACCGAGAGATCTGTGCGAGAGCCGCTGACAG 1615
DB 4520 AGAGAGCGAGCTGAGAGCTGGCGGAGAAACCGAGAGATCTGTGCGAGAGCCGCTGACAG 4579
OY 1616 TGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1675
DB 4580 TGTACTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4639
OY 1676 GGAAGCTACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
DB 4640 GGAAGCTACGAGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4699
OY 1736 TGGCGACCGGCGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1795
DB 4700 TGGAGAGGCGCGCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4759
OY 1796 TGGAGAGATGATGATCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
DB 4760 TGGAGAGATGATGATCTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4819
OY 1856 CCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1915
DB 4820 CCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4879
OY 1916 TGAACACCCCGCTGTGTGAAGCTGTGTGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAG 1975
DB 4880 TGAACACCCCGCTGTGTGAAGCTGTGTGATCAAGCTGAGAGAGAGAGAGAGAGAGAGAG 4939
OY 1976 CCGAGACCTTCTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2035
DB 4940 CCGAGACCTTCTACGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4999
OY 2036 AGGTGACCGAGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
DB 5000 AGGTGACCGAGCGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5059
OY 2096 CCGAGCTGAGGCGCATTCAGAGCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2155
DB 5060 CCGAGCTGAGGCGCATTCAGAGCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5119
OY 2156 CCGAGAGCGAGTACGCTGTGGGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
DB 5120 CCGAGAGCGAGTACGCTGTGGGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5179
OY 2216 TGTGTGAACGAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2275
DB 5180 TGTGTGAACGAGTACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5239
OY 2276 CCGCGCAAGAGGAGATCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2335
DB 5240 CCGCGCAAGAGGAGATCGGCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5299
OY 2336 GCAAGGTCTGTTCTTGAACGAGATCAATGAGGAGATGATGATCAACAGTA 2387
DB 5300 GCAAGGTCTGTTCTTGAACGAGATCAATGAGGAGATGATGATCAACAGTA 5351

```

RESULT 8
 BD263706 2312 bp DNA linear PART 17-JUL-2003
 LOCUS
 DEFINITION Improved expression of HIV polypeptides and production of

virus-like particles.
ACCESSION BD263706
VERSION BD263706.1 GI:33073474
KEYWORDS JP 2002533124-A/73.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S., Megede,U.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polyptides and production of
virus-like particles
JOURNAL Patent: JP 2002533124-A 73 08-OCT-2002;
CHIRON CORP
COMMENT OS Artificial Sequence
PN JP 2002533124-A/73
PD 08-OCT-2002
PF 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495,01-DEC-1999 US 60/168471 PI
SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI
KARIN HARTOG,
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC
C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
C12N5/10,
PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
Description of Artificial Sequence: FS(-).protomod.Rtopt(+) FH Key
FT source 1..2312
FT Location/Qualifiers
FEATURES
source : 1..2312
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
ORIGIN
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 6.1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
QY 170 GGGGAGAGAGGGCCACGATGAGATGACCGAGGCGCAACTTCTTCGGG 229
DB 1 GGGGCGCGAGAGACACCAATGAAAGATTGCACTGAGAGACAGGCTTAATTTCTTCGGG 60
QY 230 AGAAGCTGAGCTTCCCGCAGGCGCAAGCGCGCGAGTTTCCCGAGAGACGAGCA 289
DB 61 AGAGACTGAGCTTCTTCGAGAGGCGCAAGCGCGCGAGTTTCCCGAGAGACGAGCA 120
QY 290 ACAAGCCCAACAGCGCGAGCTGAGAGTGGCGCGG-----ACAACCCCGCAGGAG 343
DB 121 ACAAGCCCAACCGCGCGAGCTGAGAGTGGCGCGCGAGCAACAAGCTTGAAGGAG 180
QY 344 CCGGCGCGCAGGCGCGAGGCAACCTG-----AAGTTCCCGCAGATCAACCTGTGACG 397
DB 181 CCGGCGCGCAGCGCAGGCGCACTGAGCTTCAACTTCCCGCAGATCACTGTGACG 240
QY 398 GCCCGCTGGTGAAGCATCAAGTGGGCGCGCAAGATCAAGAGAGCGCTGTGACACCGGCG 457
DB 241 GCCCGCTGGTGAAGCATCAAGATGGGCGCGCAAGCTCAAGAGAGGCGCTGTGACACCGGCG 300
QY 458 CCGAGCAGACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATCG 517
DB 301 CCGAGCAGACCGTGTGAGAGAGATGAACTGCGCGCAAGTGAAGCCCAAGATGATCG 360
QY 518 GCGGCACTGGCGCGCTTCAATAGAGTGGCGCAAGTACAGACCAAGATCTGATGAGATCTGGG 577
DB 361 GCGGCACTGGCGCGCTTCAATAGAGTGGCGCAAGTACAGACCAAGATCTGATGAGATCTGGG 420
QY 578 GCAAGAGAGGCGCATCGGCAACGCTGTGATCGGCGCGCAACCCCGTGAACATCATCGGCGCA 637
DB 421 GCGCAAGAGGCGCATCGGCAACGCTGTGATCGGCGCGCAACCCCGTGAACATCATCGGCGCA 480
QY 638 ACAATGCTACCCAGCTGGGCTGCAACCTGAATTTCCCATCAAGCCCATCAAGACCGTGC 697

DB 481 AACTGTGACCCAGATCGGCTGACCTCTGAATTTCCCATCAAGCCCATCAAGACGCTGC 540
QY 698 CCGTGAAGCTGAAGCCCGGCAATGACGCGCGCAAGGTGAAGAGTGGCCCTTGACCGAG 757
DB 541 CCGTGAAGCTGAAGCCCGGCAATGACGCGCGCAAGGTGAAGAGTGGCCCTTGACCGAG 600
QY 758 AGAAGATCAAGGCGCTGACCGGCATCTGCGAGAGATGAGAGAGAGGCGCAAGTACCA 817
DB 601 AGAAGATCAAGGCGCTGAGATCTGCAACGAGATGAGAGAGAGGCGCAAGTACCA 660
QY 818 AGATGCGCGCGGAGAACCCCTTCAACACCCCGTGTGCGCATCAAGAGAGAGACAGCA 877
DB 661 AGATGCGCGCGGAGAACCCCTTCAACACCCCGTGTGCGCATCAAGAGAGAGACAGCA 720
QY 878 CCAAGTGGCGCAAGCTGTGAGACTTCCGAGACTGAACAGCGCACCCAGACTTCTGGG 937
DB 721 CCAAGTGGCGCAAGCTGTGAGACTTCCGAGACTGAACAGCGCACCCAGACTTCTGGG 780
QY 938 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGTGAACCGTGC 997
DB 781 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGTGAACCGTGC 840
QY 998 TGAAGTGGGCGAGCGCTTACCTTCAAGGTCGCCCTGAGACGAGACTTCCGCAAGTACACG 1057
DB 841 TGAAGTGGGCGAGCGCTTACCTTCAAGGTCGCCCTGAGACGAGACTTCCGCAAGTACACG 900
QY 1058 CCTTCAACATCCCGCAGCATCAACAAGAGACCCCGGCAATCCGCTACAGTACAGTGC 1117
DB 901 CCTTCAACATCCCGCAGCATCAACAAGAGACCCCGGCAATCCGCTACAGTACAGTGC 960
QY 1118 TGCCCCAGGGCTGGAAGGGCAGCCCCAGATTTTCCAGAGCAGCATGACCAAGATCTTGG 1177
DB 961 TGCCCCAGGGCTGGAAGGGCAGCCCCAGATTTTCCAGAGCAGCATGACCAAGATCTTGG 1020
QY 1178 AGCCCTTCGGCGCCCGGCAACCCCGAGATGTGATCTACCA-----GGCCCCCTGTAG 1231
DB 1021 AGCCCTTCGGCGAGCAAGAACCCCGAGATGTGATCTACCAAGTACAGATGACCTGTAG 1080
QY 1232 TGGGCGAGCACTGTGAGATCGGCGAGCACCGGCGCAAGATTCGAGAGTGGCGCAACCC 1291
DB 1081 TGGGCGAGCACTGTGAGATCGGCGAGCACCGGCGCAAGATTCGAGAGTGGCGCAACCC 1140
QY 1292 TGCTGGCTGGGCGCTTCAACCACTCCCGACAGAGACCAAGAGAGGCCCTTCTGT 1351
DB 1141 TGCTGGCTGGGCGCTTCAACCACTCCCGACAGAGACCAAGAGAGGCCCTTCTGT 1200
QY 1352 GGAATGGGCTACAGAGTGCACCCCGACAGTGAACGCTGAGAGCCATCGAGCTGGCGGAG 1411
DB 1201 GGAATGGGCTACAGAGTGCACCCCGACAGTGAACGCTGAGAGCCATCATGTGCGGAG 1260
QY 1412 AGGAGAGCTGAGACCGTGAACGACATTCAGAACTGTGTGGCAAGTGAATGGGCCAGCC 1471
DB 1261 AGGAGAGCTGAGACCGTGAACGACATTCAGAACTGTGTGGCAAGTGAATGGGCCAGCC 1320
QY 1472 AGATTTAACCCTGGGCAATCAAGTGTGCGCAGCTGTGCAAGTGTGTGCGCGCGCAAGGCC 1531
DB 1321 AGATTTAACCCTGGGCAATCAAGTGTGAGCACTGTGCAAGTGTGTGCGCGCGCAAGGCC 1380
QY 1532 TGAACGACATCGTGGCCCTGACCGAGAGAGCGGAGTGAAGCTGTGGCGAGAACCGGAG 1591
DB 1381 TGAACGACATCGTGGCCCTGACCGAGAGAGCGGAGTGAAGCTGTGGCGAGAACCGGAG 1440
QY 1592 TCCTGCGAGAGCCCGTGAACGAGGCTGTACTAGACCCCGAGAGAGACTGTGGCGAG 1651
DB 1441 TCCTGAGAGAGCCCGTGAACGAGGCTGTACTAGACCCCGAGAGAGACTGTGGCGAG 1500
QY 1652 TCCAGAGAGAGGCGCAACGACAGTGAACCTTACCAAGTGAAGGCGCTTCAAGAAC 1711
DB 1501 TCCAGAGAGAGGCGCAAGGCGCAGTGAACCTTACCAAGTGAAGGCGCTTCAAGAAC 1560
QY 1712 TGAAGACCGGCAAGTACCGCAAGTGGCAACCGGCCCAACCAACGACGTAAGACACTGA 1771

Db 1561 TGAAGACCGGACGATACGCCCCGATGCGGCGGCCCAACACGACGTGAAGACCTGA 1620
 QY 1772 CCGAGGCGCTGACGAGATGCGCATGAGAGCATGTCGTGGGCAAGACCCCAAGT 1831
 Db 1621 CCGAGGCGCTGACGAGATGCGCATGTCGTGGGCAAGACCCCAAGT 1680
 QY 1832 TCCGCTGCGCATCCAGAGAGAGACCTGAGAGACCTGTGAGACCGATCTGAGAGGCA 1891
 Db 1681 TCAAGCTGCGCATCCAGAGAGAGACCTGAGAGGCGCTGTGAGATGTCGTGGAGGCA 1740
 QY 1892 CCGTGAATCCCGAGTGGAGAGTTCGTGAACACCCCGCTGTGTAAGCTGTGTACAGC 1951
 Db 1741 CCGTGAATCCCGAGTGGAGAGTTCGTGAACACCCCGCTGTGTAAGCTGTGTACAGC 1800
 QY 1952 TGAAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGACCGGCGCGCAACCGCG 2011
 Db 1801 TGAAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTGAGACCGGCGCGCAACCGCG 1860
 QY 2012 AGACCAAGATCGGCAAGGCGCGCTACGTGACCGACCGGCGCGCGAGAGATCGTAGGCC 2071
 Db 1861 AGACCAAGATCGGCAAGGCGCGCTACGTGACCGACCGGCGCGCGAGAGATCGTAGGCC 1920
 QY 2072 TGAACGAGACCAACCAACCAAGAGACCTGAGCTGAGAGCTGAGCTGCGTGGAGAGCA 2131
 Db 1921 TGAACGAGACCAACCAACCAAGAGACCTGAGCTGAGAGCTGAGCTGCGTGGAGAGCA 1980
 QY 2132 GCGGACGAGAGTGAACATCGTGAACGACGACGACGACCTGAGGACATCATCAAGGCGC 2191
 Db 1981 GCGGACGAGAGTGAACATCGTGAACGACGACGACGACCTGAGGACATCATCAAGGCGC 2040
 QY 2192 AGCCCGACAAAGCGAGACGAGCTGTGAACCAAGTCAATCAAGACCTGATCAAGAGG 2251
 Db 2041 AGCCCGACAAAGCGAGACGAGCTGTGAACCAAGTCAATCAAGACGCTGATCAAGAGAG 2100
 QY 2252 AGAAGGTGACCTGAGCTGGGTGCGCCCAACAGGCGCATCGCGCGCAACGAGCAGATG 2311
 Db 2101 AGAAGGTGACCTGAGCTGGGTGCGCCCAACAGGCGCATCGCGCGCAACGAGCAGATG 2160
 QY 2312 ACAAGCTGTGAGACGAGGCGCATCGGCAAGTGTCTGTGACGCGGCAATCGATGAGCGCA 2371
 Db 2161 ACAAGCTGTGAGACGAGGCGCATCGGCAAGTGTCTGTGACGCGGCAATCGATGAGCGCA 2220
 QY 2372 TCGTATCTACAGTACATGAGACGACCTGTACGTGAGCGCGCGCGCTAGAGATCATT 2431
 Db 2221 TCGTATCTACAGTACATGAGACGACCTGTACGTGAGCGCGCGCGCTAGAGATCATT 2280
 QY 2432 AAAAGCTTCCCGGCGCTAGACCGGTTGAATTC 2463
 Db 2281 AAAAGCTTCCCGGCGCTAGACCGGTTGAATTC 2312

RESULT 9
 LOCUS CQ870576 2312 bp DNA linear PAT 13-SBP-2004
 DEFINITION Sequence 84 from Patent EP1433851.
 ACCESSION CQ870576
 VERSION CQ870576.1 GI:52000092
 KEYWORDS
 SOURCE
 ORGANISM
 1
 other sequences; artificial sequences.
 REFERENCE
 AUTHORS Barnett,S., Zumedege,J., Srivastava,T., Lian,Y., Hartog,K.,
 Liu,H., Greer,C., Selby,M. and Walker,C.
 TITLE Improved expression of HIV polypeptides and production of
 JOURNAL virus-like particles
 CHIRON CORPORATION (US)
 FEATURES
 source 1..2312
 location/Qualifiers
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"

ORIGIN
 /note="Description of Artificial Sequence:
 FS (-).promod.Rlopt(+)"

Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
 Best Local Similarity 93.3%; Pred. No. 6.1e-165;
 Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGCGCCACGAGATGAAGGACTGACCGAGCGGCCCAAGCTTCTCCGCG 229
 Db 1 GCGGCGCGAGAGACCAATGAAGATGACATGAGAGACGCTAATTTCTTCGCG 60
 QY 230 AGGACCTGCGCTTCCCGCAAGGCGCGCGAGATTCCCGAGAGAGCAACCGCGCA 289
 Db 61 AGGACCTGCGCTTCCCGCAAGGCGCGCGAGATTCCCGAGAGAGCAACCGCGCA 120
 QY 290 ACAAGCCCAACAGCCCGAGACTGCAAGTGCAGCGGCGG-----ACAAGCCCGAGAGAG 343
 Db 121 ACAAGCCCAACAGCCCGAGACTGCAAGTGCAGCGGCGGCGGAGAACAGAGCTGAGCGAGG 180
 QY 344 CCGGCGCGAGGCGCGAGGCGACCCCTG-----AATTCGCCAGATCAACCTGTGGCAGC 397
 Db 181 CCGGCGCGAGCGCGAGGCGACCGTGAAGCTTCACTTCCCGAGATCAACCTGTGGCAGC 240
 QY 398 GCGCCCTGTGAGCATCAAGTGGCGCGCGAGATCAAGAGGCGCTGTGAGCAACCGCG 457
 Db 241 GCGCCCTGTGAGCATCAAGATCGGCGCGAGCTCAAGAGGCGCTGTGAGCAACCGCG 300
 QY 458 CCGATGACACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGGCCCAAGATGATG 517
 Db 301 CCGAGACACCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGGCCCAAGATGATG 360
 QY 518 GCGGACGAGGCGGCTTCAATCAAGTGGCGAGTACGACCAAGATCTGTATCGAGATCTG 577
 Db 361 GCGGACGAGGCGGCTTCAATCAAGTGGCGAGTACGACCAAGATCTGTATCGAGATCTG 420
 QY 578 GCAAGAGGCGCATCGGACCGTGTGATCGGCGCGCACCCCGTGAACATCATCGGCGCA 637
 Db 421 GCGCAAGGCGCATCGGACCGTGTGATCGGCGCGCACCCCGTGAACATCATCGGCGCA 480
 QY 638 ACATGCTGACCGAGCTGGCTGACCGCTGAATTTCCCATCAAGCCCGCATGAGACCGTGC 697
 Db 481 ACCTGCTGACCGAGATCGGCTGACCGCTGAATTTCCCATCAAGCCCGCATGAGACCGTGC 540
 QY 698 CCGTGAAGCTGAAGCGCGCGAGTGAAGCGCGCGCAAGTGAAGAGTGGCGCCCTGAACGAG 757
 Db 541 CCGTGAAGCTGAAGCGCGCGAGTGAAGCGCGCGCAAGTGAAGAGTGGCGCCCTGAACGAG 600
 QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGCGCAAGATCAACA 817
 Db 601 AGAAGATCAAGGCGCTGAGTGAAGATCTGACCGAGATGAGAGAGGCGCAAGATCAACA 660
 QY 818 AGATGCGCCCGAGAACCCCTTACAAACCCCGTGTTCGCATCAAGAGAGAGACAGCA 877
 Db 661 AGATGCGCCCGAGAACCCCTTACAAACCCCGTGTTCGCATCAAGAGAGAGACAGCA 720
 QY 878 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGACGTGAACAAGGCGACCCGAGACTTCTGG 937
 Db 721 CCAAGTGGCGCAAGCTGTGAGCTTCCGAGACGTGAACAAGGCGACCCGAGACTTCTGG 780
 QY 938 AGGTGAGCTGGAGCATCCCGCACCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 997
 Db 781 AGGTGAGCTGGAGCATCCCGCACCGCGCGCGCTGAAGAGAGAGAGAGAGAGAGAGAG 840
 QY 998 TGAAGTGGAGAGAGCTTACCTTACAGGTGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 1057
 Db 841 TGAAGTGGAGAGAGCTTACCTTACAGGTGCGCTTGAAGAGAGAGAGAGAGAGAGAGAG 900
 QY 1058 CTTTACCAATCCCGAGAGCATCAACAAGAGACCCCGCGCATTCGCTACAGATCAAGTGC 1117
 Db 901 CTTTACCAATCCCGAGAGCATCAACAAGAGACCCCGCGCATTCGCTACAGATCAAGTGC 960
 QY 1118 TGCAGGAGGCTGAGAGGCGAGCGCGAGCATCTTCAAGAGAGCATGACCAAGATCTG 1177

Db 961 TCCCCAGGAGGCTGGAAGGAGAGCCCGCCATCTTCCAGAGCAGATGACCAAGATCTG 1020
Qy 1178 AGCCCTTCGCGCCGCGCAACCCCGAGATCGTATCTACCA-----GACCCCTCTGAGC 1231
Db 1021 AGCCCTTCGCGCAAGAGAGAGCCCGCATCGTATCTACCAAGTACATGAGAGAGCTGAGC 1080
Qy 1232 TGGGAGGCACTGAGATCGGCGAGACCGCGCCAAAGTTCAGAGAGCTGCGCAAGACCC 1291
Db 1081 TGGGAGGCACTGAGATCGGCGAGACCGCAACCAAGATCGAGAGCTGCGCGCAAGACCC 1140
Qy 1292 TGGTGGCTGGGGCTTCAACACCCCGCAAGAGAGCAAGAGAGAGCCCTTCTG 1351
Db 1141 TGTGGCTGGGGCTTCAACACCCCGCAAGAGAGCAAGAGAGAGCCCTTCTG 1200
Qy 1352 GGATGGGCTACAGAGCTGCAACCCCGCAAGTGAACCTGTGAGCCCATGAGCTGCGCGAG 1411
Db 1201 GGATGGGCTACAGAGCTGCAACCCCGCAAGTGAACCTGTGAGCCCATGAGCTGCGCGAG 1260
Qy 1412 AGGAGAGCTGAGACCTGGAACGATCTCAGAAAGCTGGTGGCAAGCTGAACTGGGCGAGC 1471
Db 1261 AGGAGAGCTGAGACCTGGAACGATCTCAGAAAGCTGGTGGCAAGCTGAACTGGGCGAGC 1320
Qy 1472 AGATCTACCCCGGAGCAAGAGTGGCGAGCTGTGCAAGCTGTGGCGGCGCGCAAGGCGC 1531
Db 1321 AGATCTACCCCGGAGCAAGAGTGGCGAGCTGTGCAAGCTGTGGCGGCGCGCAAGGCGC 1380
Qy 1532 TGAACGACATCGTACCCTGACCCGAGAGAGCCGAGCTGAGAGCTGGCCGAGAACCGCGAG 1591
Db 1381 TGAACGAGGTGATCCCCCTGACCGAGAGAGCCGAGCTGAGAGCTGGCCGAGAACCGCGAG 1440
Qy 1592 TCCGCGCGAGAGCCCTGTCAGAGCGGTGTACTAGACCCCGAGAGAGAGCTGTGGCGAG 1651
Db 1441 TCCGAGAGAGAGCCCTGTCAGAGGTGTACTAGACCCCGAGAGAGAGCTGTGGCGAG 1500
Qy 1652 TCCAGAGAGAGAGCCGAGACAGTGAACCTACCAAGATCTACCGAGAGCCCTTCAAGAAC 1711
Db 1501 TCCAGAGAGAGAGCCGAGACAGTGAACCTACCAAGATCTACCGAGAGCCCTTCAAGAAC 1560
Qy 1712 TGAAGACCGGCAAGTACGCAAGATGCGACCGCCCAACCAAGAGTGAAGCAGCTGA 1771
Db 1561 TGAAGACCGGCAAGTACGCGCGATGCGCGCCCGCCCAACCAAGAGTGAAGCAGCTGA 1620
Qy 1772 CCGAGGCGCTGCAAGAGATCGCCATGAGAGAGATCTGTATCTGGGCGAAGACCCCAAGT 1831
Db 1621 CCGAGGCGCTGCAAGAGTGAACCGAGAGAGATCTGTATCTGGGCGAAGATCCCAAGT 1680
Qy 1832 TCGGCTGCGCCATCTCAGAAAGAGAGCTGGGAGAGCTGTGTGAGACCGACTACTGGCAGGCA 1891
Db 1681 TCAAGCTGCGCCATCTCAGAAAGAGAGCTGGGAGAGCTGTGTGAGATGAGTACTGGCAGGCA 1740
Qy 1892 CCGTGAATCCCGAGTGGAGATTGTGTGAACACCCCGCCCTGTGAAAGCTGTGTGATCAGC 1951
Db 1741 CCGTGAATCCCGAGTGGAGATTGTGTGAACACCCCGCCCTGTGTGAAAGCTGTGTGATCAGC 1800
Qy 1952 TGAAGAGAGAGCCCATCTCGCGCGCGAGACCTTCTACGTGAGACGCGCGCCGCAACCGCG 2011
Db 1801 TGAAGAGAGAGCCCATCTGTGGCGCGAGACCTTCTACGTGAGACGCGCGCCGCAACCGCG 1860
Qy 2012 AGACCAAGATCGGCAAGAGCGCGCTACGTGAACGACCGGGGCGCGCAAGAAATGTGTAGCC 2071
Db 1861 AGACCAAGATCGGCAAGAGCGCGCTACGTGAACGACCGGGGCGCGCAAGAAATGTGTAGCA 1920
Qy 2072 TGAACGAGACCAACCAAGAGAGCGAGCTGAGAGCCATCTCAGTGGCCCTGCAAGAGCA 2131
Db 1921 TGCAGGACACCAACCAAGAGAGCGAGCTGAGAGCCATCTCAGTGGCCCTGCAAGAGCA 1980
Qy 2132 GCGGAGAGAGTGAACATCTGTGACCGACAGCCAGTACGCGCTGTGGAGTATCTCAGGCGC 2191
Db 1981 GCGGAGAGAGTGAACATCTGTGACCGACAGCCAGTACGCGCTGTGGAGTATCTCAGGCGC 2040
Qy 2192 AGCCCAAGAGAGAGAGAGCTGTGAAACCAAGATCTCAGAGAGCTGTATCAAGAGG 2251

Db 2041 AGCCCGACAGAGAGCGAGAGCGAGCTGTGTGAGCCAGATCATCGAGAGCTGATCAAGAGG 2100
Qy 2252 AGAAGGTGATCTGAGCTGGGTGCGCCCGCCCAAGAGGATTCGGCGGCAACGAGCAGATGG 2311
Db 2101 AGAAGGTGATCTGAGCTGGGTGCGCCCGCCCAAGAGGATTCGGCGGCAACGAGCAGATGG 2160
Qy 2312 ACAAGCTGTGAGCAAGAGGATCCGCAAGGTGTGTCTGTGAGCGGCAATCATGTGGCGCA 2371
Db 2161 ACAAGCTGTGTGAGCGCGGATTCGCAAGGTGTGTCTGTGAAACGAGCATGATGCGGCA 2220
Qy 2372 TGTGATCTACCAAGTACATGAGAGCACTGTACGTGTGGGAGCGCGCCCTAGAGTGAAT 2431
Db 2221 TGTGATCTACCAAGTACATGAGAGCACTGTACGTGTGGGAGCGCGCCCTAGAGTGAAT 2280
Qy 2432 AAAAGCTTCCGCGGCTGACACCGGTGAATTC 2463
Db 2281 AAAAGCTTCCGCGGCTGACACCGGTGAATTC 2312

RESULT 10
AR373389
LOCUS AR373389 2312 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLES Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
Chiron Corporation; Emeryville, CA
LOCATION/Qualifiers
FEATURES
source 1..2312
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 82.2%; Score 2025.2; DB 6; Length 2312;
Best Local Similarity 93.3%; Pred. No. 6,1e-165;
Matches 2156; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
Qy 170 GCGGCAAGAGAGGCGCAACAGATGAGAGCTGCAACGAGGCGCAGGCCAACTTCTTCGCG 229
Db 1 GCGGCGGCGAAGAGACCAAGATGAGATGAGAGCTGAGAGAGAGCTTCTTCGCG 60
Qy 230 AGGACTGAGCTTCCCGAGAGGCGAGGCGCGAGTTCCCGAGAGAGAGAGAGCGCGCA 289
Db 61 AGGACTGAGCTTCCCGAGAGGCGAGGCGCGAGTTCCCGAGAGAGAGAGAGCGCGCA 120
Qy 290 ACAAGCCCAACAGCCCGAGAGCTGCAAGTGGCGCG-----ACAACCCCGGAGAGG 343
Db 121 ACAAGCCCAACAGCCCGAGAGCTGCAAGTGGCGCGCGCGAGAACCAAGCTGAGAGG 180
Qy 344 CCGGCGCGAGGCGCAGAGGCAACCTG-----AACTTCCCGAGATCAACCTGTGAGAGC 397
Db 181 CCGGCGCGAGGCGCAGAGGCAACCTGAGCTTCAACTTCCCGAGATCAACCTGTGAGAGC 240
Qy 398 GCCCTGTGTGAGATCAAGTGGCGCGCGAGTCAAGAGAGCGCTGTGAGACCGCGC 457
Db 241 GCCCTGTGTGAGATCAAGTGGCGCGCGAGTCAAGAGAGCGCTGTGAGACCGCGC 300
Qy 458 CCGAGACACCGGTGTGAGAGAGTGAAGCTGCGCGGCAAGTGAAGGCCCAAGATGATG 517
Db 301 CCGAGACACCGGTGTGAGAGAGTGAAGCTGCGCGGCAAGTGAAGGCCCAAGATGATG 360
Qy 518 GCGGAGTGGCGGCTTCAATCAAGTGGCGCGAGTACAGACAGATCTGATGAGATCTGCG 577
Db 361 GCGGAGTGGCGGCTTCAATCAAGTGGCGCGAGTACAGACAGATCCCGGTGAGATCTGCG 420

QY 578 GCAGAGAGGCATCGGCAACGTCGTGATCGGCCCAACCCCGTGAACATCATCGGCCGA 637
 DB 421 GCCAAGAGGCATCGGCAACGTCGTGATCGGCCCAACCCCGTGAACATCATCGGCCGA 480
 QY 638 ACATGTCGACCCAGGCTGGCTGCACTTGAATCCCATCAAGCCCATCGAGACCGTGC 697
 DB 481 ACCTGTGACCCAGGATCGGCTGCACTTGAATCCCATCAAGCCCATCGAGACCGTGC 540
 QY 698 CGGTGAAGCTGAAGCCCGGCAATGAGACGCGCCCAAGGATGAGAGTGGCCCTGACCGAG 757
 DB 541 CGGTGAAGCTGAAGCCCGGGAATGAGACGCGCCCAAGGATGAGAGTGGCCCTGACCGAG 600
 QY 758 AGAAGATCAAGGACCTTGACCGCATCTCGAGAGAGATGAGAGAGGCGCAAGATCA 817
 DB 601 AGAAGATCAAGGACCTTGATGAGATCTGACCGAGATGAGAGAGGCGCAAGATCA 660
 QY 818 AGATGGGCGCCAGAGAACCCCTTCAACACCCCGTGTTCGCCATCAAGAGAGAGACAGCA 877
 DB 661 AGATGGGCGCCAGAGAACCCCTTCAACACCCCGTGTTCGCCATCAAGAGAGAGACAGCA 720
 QY 878 CCAAGTGGCGCAAGCTGTGGACTTCCGAGAGCTGAACAGGCGCACCCGAGCTTCTGG 937
 DB 721 CCAAGTGGCGCAAGCTGTGGACTTCCGAGAGCTGAACAGGCGCACCCGAGCTTCTGG 780
 QY 938 AGGTGAGTGGGCAATCCCGCAACCCCGCGCGCTGAAGAGAGAGAGAGCGTGAACCGTGC 997
 DB 781 AGGTGAGTGGGCAATCCCGCAACCCCGCGCGCTGAAGAGAGAGAGAGCGTGAACCGTGC 840
 QY 998 TGGAGTGGGCGACCGCTTCAAGCTGTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1057
 DB 841 TGGAGTGGGCGACCGCTTCAAGCTGTGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 QY 1058 CCTTCACATCCCGAGCATCAACAGAGACCCCGGCGCATCCGCTACAGTCAACAGTGC 1117
 DB 901 CCTTCACATCCCGAGCATCAACAGAGACCCCGGCGCATCCGCTACAGTCAACAGTGC 960
 QY 1118 TGGCGGAGGCTGAGAGAGGCAAGCCCGAGCATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1177
 DB 961 TGGCGGAGGCTGAGAGAGGCAAGCCCGAGCATCTTCAAGAGAGAGAGAGAGAGAGAGAG 1020
 QY 1178 AGCCCTTCGCGCGCGCGCAACCCCGAGATGTGATCTACCA-----GGCCCGCTGTAGC 1231
 DB 1021 AGCCCTTCGCGAG 1080
 QY 1232 TGGGAGAGGCACTGAGAGATCGGCGAGCAACGCGGCAAGATCGAGAGAGAGAGAGAGAG 1291
 DB 1081 TGGGAGAGGCACTGAGAGATCGGCGAGCAACGCGGCAAGATCGAGAGAGAGAGAGAGAG 1140
 QY 1292 TGGTGGCTGGGAGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1351
 DB 1141 TGGTGGCTGGGAGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 QY 1352 GGATGGGCTGAGAGAGTGCACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1411
 DB 1201 GGATGGGCTGAGAGAGTGCACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
 QY 1412 AGGAGAGCTGAGACCGTGAAGAGATCTCAAGAGAGTGGTGGGCAAGCTGAGAGAGAGAG 1471
 DB 1261 AGGAGAGCTGAGACCGTGAAGAGATCTCAAGAGAGTGGTGGGCAAGCTGAGAGAGAGAG 1320
 QY 1472 AGATCTACCCCGGCACTCAAGAGTGGCGAGAGTGTGCAAGCTGTGCGCGCGCAAGAGAG 1531
 DB 1321 AGATCTACCCCGGCACTCAAGAGTGGCGAGAGTGTGCAAGCTGTGCGCGCGCAAGAGAGAG 1380
 QY 1532 TGAACGACATCTGTGCTTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1591
 DB 1381 TGAACGAGGTGATCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
 QY 1592 TCTGTGGAG 1651
 DB 1441 TCTGTGGAG 1500
 QY 1652 TCCAG 1711

DB 1501 TCCAG 1560
 QY 1772 TGAAG 1771
 DB 1561 TGAAG 1620
 QY 1772 CCGAG 1831
 DB 1621 CCGAG 1680
 QY 1832 TCCGCTTGGCCATCCAG 1891
 DB 1681 TCAAGCTGCGCATCCAG 1740
 QY 1892 CCGTGAATCCCGAG 1951
 DB 1741 CCGTGAATCCCGAG 1800
 QY 1952 TGGAG 2011
 DB 1801 TGGAG 1860
 QY 2012 AGACCAAGATCGGCAAG 2071
 DB 1861 AGACCAAGATCGGCAAG 1920
 QY 2072 TGAACGAG 2131
 DB 1921 TGAACGAG 1980
 QY 2132 GCGGAG 2191
 DB 1981 GCGGAG 2040
 QY 2192 AGCCGAG 2251
 DB 2041 AGCCGAG 2100
 QY 2252 AGAAGGTGATCTGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2311
 DB 2101 AGAAGGTGATCTGAGCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2312 ACAAGCTGTGAG 2371
 DB 2161 ACAAGCTGTGAG 2220
 QY 2372 TCGTATCTACCAAGTGAATGAG 2431
 DB 2221 TCGTATCTACCAAGTGAATGAG 2280
 QY 2432 AAAAGCTTCCGAG 2463
 DB 2281 AAAAGCTTCCGAG 2312

RESULT 11
 BD263705 2300 bp DNA linear PART 17-UTL-2003
 LOCUS Improved expression of HIV polypeptides and production of
 DEFINITION virus-like particles.
 ACCESSION BD263705
 VERSION BD263705.1 GI:33073473
 KEYWORDS UP 200253124-A/72.
 SOURCE synthetic construct
 ORGANISM other sequences; artificial sequences.
 REFERENCE 1 (bases 1 to 2300)
 AUTHORS Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
 Greer,C., Selby,M. and Walker,C.
 TITLE Improved expression of HIV polypeptides and production of
 virus-like particles
 JOURNAL Patent: JP 200253124-A 72 08-OCT-2002;

COMMENT CHIRON CORP
OS Artificial Sequence
PN JP 200253124-A/72
PD 08-OCT-2002
PR 30-DEC-1999 JP 2000591193
PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471 PI
SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIYASTAVA, YING LIAN, PI
KARIN HARTOG,
PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC
C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
C12N5/10,
PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
Description of Artificial Sequence: PS(-). protomod. Rtopc. YMM FM
Key Location/Qualifiers
FT source 1..2300
/organism='Artificial Sequence'.
FEATURES
source Location/Qualifiers
1..2300
/organism='synthetic construct'
/mol_type='genomic DNA'
/db_xref='taxon:32630'

ORIGIN
Query Match 82.0%; Score 2019.2; DB 6; Length 2300;
Best Local Similarity 93.2%; Pred. No. 2e-164;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGGCCACGATGAGAGCTGCAACGAGCGCCAGCCAACTTCTCCGCG 229
DB 1 GCGGCGCGAAGAGCAACCAATGAAAGATTGCACTGAGAGACAGGCTAATTCTTCCGCG 60
QY 230 AGGACTGCGCTTCCCAAGGGCAAGCCCGCGAGTTCCCAAGAGCAAGAACCGCGCA 289
DB 61 AGGACTGCGCTTCTGCAAGGGCAAGCCCGCGAGTTCAAGAGCAAGAACCGCGCA 120
QY 290 ACAAGCCCCACAGCCCGGAGCTGCAAGTGGCGCG-----ACAACCCCCGAGAGAG 343
DB 121 ACAAGCCCCACAGCCCGGAGCTGCAAGTGGCGCGAGAGCAACAGCTGAGGAG 180
QY 344 CCGGCGCGAGGGCCAGGGGCAACCTG-----AACTTCCCGAGATCAACCTTGGCAG 397
DB 181 CCGGCGCGAGGGCCAGGGGCAACCTGAACTTCACTTCCCGAGATCAACCTTGGCAG 240
QY 398 GCCCCTGCTGAGCATCAAGTGGCGCGCCAGATCAAGAGCCCTGCTGAGCAACCGCG 457
DB 241 GCCCCTGCTGAGCATCAAGTGGCGCGCCAGCTCAAGAGGCGCTGCTGAGCAACCGCG 300
QY 458 CCGAGCAACCTGCTGAGAGAGATGAGCTGCCCGGCAAGTGAAGCCCAATGATGCG 517
DB 301 CCGAGCAACCTGCTGAGAGAGATGAACTGCGCGGCAAGTGAAGCCCAATGATGCG 360
QY 518 GCGGCAACCGGCGCTTCAATGAGTGGCGCCAGTACGACAGATCTGATCGAGATGCG 577
DB 361 GCGGCAACCGGCGCTTCAATGAGTGGCGCGAGTACGACAGATCTGATCGAGATGCG 420
QY 578 GCAAGAGGCGCATCGGCAACCTGCTGATCGGCCCAACCCCGTGAACATCATCGCGCA 637
DB 421 GCAAGAGGCGCATCGGCAACCTGCTGATCGGCCCAACCCCGTGAACATCATCGCGCA 480
QY 638 ACATGCTGAGCCAGCTGGGGCTGCAACCTTGAACCTTCCCATCAAGCCCATGAGACGCTGC 697
DB 481 ACCTGCTGAGCCAGATCGGCTGCAACCTTGAACCTTCCCATCAAGCCCATGAGACGCTGC 540
QY 698 CCGTGAAGCTGAAGCCCGGATGAGAGCGCCCAAGGTGAAGAGAGTGGCCCTTGAACGAG 757
DB 541 CCGTGAAGCTGAAGCCCGGATGAGAGCGCCCAAGGTGAAGAGAGTGGCCCTTGAACGAG 600
QY 758 AGAAGATCAAGGCGCTGAGCCGCACTTGCAGAGAGATGAGAGAGAGGCGCAAGATCA 817
DB 601 AGAAGATCAAGGCGCTTGTGAGATCTGCAACGAGATGAGAGAGAGGCGCAAGATCA 660
QY 818 AGATCGGCGCGAGAACCCCTTACCAACCCCGCTGTTGCCATTCAGAGAGAGAGCA 877

DB 661 AGATCGGCGCGAGAACCCCTTACCAACCCCGCTGTTGCCATTCAGAGAGAGAGCA 720
QY 878 CCAAGTGGCGCAAGCTGCTGAGCTTCCGAGAGCTGAACAAGCGCACCCAGAGACTTGGG 937
DB 721 CCAAGTGGCGCAAGCTGCTGAGCTTCCGAGAGCTGAACAAGCGCACCCAGAGACTTGGG 780
QY 938 AGGTGAGTGGCGCATCCCGGAGCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAG 997
DB 781 AGGTGAGTGGCGCATCCCGGAGCGCGCGGCTGAAGAGAGAGAGAGAGAGAGAGAG 840
QY 998 TGAAGTGGCGCAAGCGCTTCACTTCAAGCTGCGCCCTGAGCAAGAGACTTCCGAGTACCG 1057
DB 841 TGAAGTGGCGCAAGCGCTTCACTTCAAGCTGCGCCCTGAGCAAGAGACTTCCGAGTACCG 900
QY 1058 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCGCATCCGCTTCAAGTACAGTGC 1117
DB 901 CCTTCAACATCCCGAGCATCAACAGAGACCCCGGCGCATCCGCTTCAAGTACAGTGC 960
QY 1118 TGGCCAGAGGCTGAGAGAGGCGAGCCCGAGAGATGATCTACAGAGAGAGAGAGAGAG 1177
DB 961 TGGCCAGAGGCTGAGAGAGGCGAGCCCGAGAGATGATCTACAGAGAGAGAGAGAGAG 1020
QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGAGTGTATCTACAGAGGCGCCCTGTACGTGGGCA 1237
DB 1021 AGCCCTTCCGCGCGCGCAACCCCGAGAGTGTATCTACAGAGGCGCCCTGTACGTGGGCA 1080
QY 1238 GCGACTGAGATCGGCGAGCAACCGGCGCAAGTGAAGAGAGTGGCGCAAGCCTGCTGC 1297
DB 1081 GCGACTGAGATCGGCGAGCAACCGGCGCAAGATGAAGAGAGTGGCGCGCAGCCTGCTGC 1140
QY 1298 GCTGGGAGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
DB 1141 GCTGGGAGCTTCAACACCCCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1198
QY 1358 GCTACAGAGTGAACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
DB 1199 ---CGAGCTGAGACCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1254
QY 1418 GCTGAGACGCTGAACGACATCAAGAGAGTGTGGCAAGCTGAACCTGGGCGAGCAGATCT 1477
DB 1255 GCTGAGACGCTGAACGACATCAAGAGAGTGTGGCAAGCTGAACCTGGGCGAGCAGATCT 1314
QY 1478 ACCCGGCGCATCAAGTGGCGCGAGCTGTGCAAGTGTGGCGCGCGCGCAAGCCCTGAC 1537
DB 1315 ACCCGGCGCATCAAGTGGCGCGAGCTGTGCAAGTGTGGCGCGCGCGCAAGCCCTGAC 1374
QY 1538 ACATGTGCGCCCTGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
DB 1375 AGGTGATCCCTTGAACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1434
QY 1598 GCGAGCCCGTGAACGAGCGTGTACTACAGACCCCGAGAGAGAGAGAGAGAGAGAG 1657
DB 1435 AGGAGCCCGTGAACGAGCGTGTACTACAGACCCCGAGAGAGAGAGAGAGAGAGAG 1494
QY 1658 AGGAGGCGCAACGAGTGAAGCTTACAGAGATCTTACAGAGAGAGAGAGAGAGAG 1717
DB 1495 AGGAGGCGCAACGAGTGAAGCTTACAGAGATCTTACAGAGAGAGAGAGAGAGAG 1554
QY 1718 CCGGCAAGTACGCGCAAGATGGCGCAACGAGAGAGAGAGAGAGAGAGAGAGAGAG 1777
DB 1555 CCGGCAAGTACGCGCGAGATGGCGCGAGCGCGCGAGAGAGAGAGAGAGAGAGAG 1614
QY 1778 CCGTGAAGAGATCGCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1837
DB 1615 CCGTGAAG 1674
QY 1838 TGGCCATCAAG 1897
DB 1675 TGGCCATCAAG 1734
QY 1898 TGGCCAG 1957
DB 1735 TGGCCAG 1794

Db 1141 GCTGGGGCTTACACACCCCCGACAAAGACACAGAGAGGCCCTTCTGCCAT-- 1198
Qy 1358 GCTAGAGCTGACCCCGACAAAGTGAACCTGACAGCCCATGAGCTGCCGGAAGAGAGA 1417
Db 1199 ----CGAGCTGACCCCGACAAAGTGAACCTGACAGCCCATGAGCTGCCGGAAGAGAGA 1254
Qy 1418 GCTGACCGTGAACACATCCGAAAGCTGTGGGCAAGCTGAACCTGGGCAACCCAGATCT 1477
Db 1255 GCTGACCGTGAACACATCCGAAAGCTGTGGGCAAGCTGAACCTGGGCAACCCAGATCT 1314
Qy 1478 ACCCGGACATCAAGGTGCGCCAGCTGTGCAAGCTGTGCGCGCCGCAAGGCCCTGACCG 1537
Db 1315 AGCGGACATCAAGGTGAGAGAGAGCTGTGCAAGCTGTGCGCGCCGCAAGGCCCTGACCG 1374
Qy 1538 ACATGTCGCTCTGACCGAGAGAGCCGAGCTGAGAGTGGCCGAGAACCGCGAGATCTGCG 1597
Db 1375 AGGTATATCCCTTGACCGAGAGAGCCGAGCTGAGAGTGGCCGAGAACCGCGAGATCTGTA 1434
Qy 1598 GCGAGCCCTGACCGCGCTGTACTAGAACCCGACGAAGACCTGTGTGGCCGAGATCTGTA 1657
Db 1435 AGGAGCCCTGACCGAGGTGTACTAGAACCCGACGAAGACCTGTGTGGCCGAGATCTGTA 1494
Qy 1658 AGCAGGCGCCAGACCGAGTGAACCTGACAGATCTGACGAGAGCCCTTCAAGAACCTGAGA 1717
Db 1495 AGCAGGCGCCAGAGCCAGTGAACCTGACAGATCTGACGAGAGCCCTTCAAGAACCTGAGA 1554
Qy 1718 CCGGCAAGTACGCAAGATGCGCACCGCCACACCAACGATGAAGAGCTGACCGAGG 1777
Db 1555 CCGGCAAGTACGCGCGCATGCGCGCGCCACACCAACGATGAAGAGCTGACCGAGG 1614
Qy 1778 CCGTGAGAAAGTCCGATGAGAGAGATGTGTATCTGGGCGCAAGACCCCGCAAGTTCCGCG 1837
Db 1615 CCGTGAGAAAGTGAACCGAGAGATGTGTATCTGGGCGCAAGATCTCCCAAGTTCAAGC 1674
Qy 1838 TGGCCATCCAGAGAGAGACCTGGGAGACCTGTGTGACCGACTGAGGAGCCGACCTGGA 1897
Db 1675 TGGCCATCCAGAGAGAGACCTGGGAGACCTGTGTGAGTGAAGTGTGAGAGCCGACCTGGA 1734
Qy 1898 TCCCGGAGTGGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGAGA 1957
Db 1735 TCCCGGAGTGGAGTTCGTGAACACCCCGCTGTGTGAGCTGTGTGATCCAGCTGAGA 1794
Qy 1958 AGGAGCCCATCATCGCGCGCGAGACCTTCTGACTGAGACGCGCGCCGCAACCGCGAGACA 2017
Db 1795 AGGAGCCCATCATCGTGGCGCGCGAGACCTTCTGACTGAGACGCGCGCCGCAACCGCGAGACA 1854
Qy 2018 AGATGGGCAAGGCGGCTGATGTAACGACCGGGGCGGGGAGAGATGTAAGCTGACCG 2077
Db 1855 AGCTGGGCAAGGCGGCTGATGTAACGACCGGGGCGGGGAGAGATGTAAGCTGACCG 1914
Qy 2078 AGACCAACCAAGAGACCGAGCTGACAGCCATCCAGCTGCGCTGACAGACAGCGGGA 2137
Db 1915 AGACCAACCAAGAGACCGAGCTGACAGCCATCCAGCTGCGCTGACAGACAGCGGCG 1974
Qy 2138 GCGAGGTGAACATGTGACCGACGACGACGACCTTGGGCAATCCAGGCCGACCGCG 2197
Db 1975 TGGAGGTGAACATGTGACCGACGACGACGACCTTGGGCAATCCAGGCCGACCGCG 2034
Qy 2198 ACAGAGCGGAGAGCGAGCTGTGTGAACCAATCATTCAGAGCTGATCCAGAGAGAGAG 2257
Db 2035 ACAGAGCGGAGAGCGAGCTGTGTGAACCAATCATTCAGAGCTGATCCAGAGAGAGAG 2094
Qy 2258 TGTACTGAGCTGGTGGCGCCGACCAAGAGGCAATCGCGCGCAACGAGAGATGACAAAG 2317
Db 2095 TGTACTGAGCTGGTGGCGCCGACCAAGAGGCAATCGCGCGCAACGAGAGATGACAAAG 2154
Qy 2318 TGTGAGCAAGGCAATCCGCAAGGTGTCTTCTGACCGGCAATGATGCGCGCATGTGTA 2377
Db 2155 TGTGAGCGCGCGCATTCGCAAGGTGTCTTCTGACCGGCAATGATGCGCGCATGTGTA 2214
Qy 2378 TCTACAGATCAAGAGCACTGTGACGTGGGAGCGCGCGCTAGAGATGATTAAAGC 2437
Db 2215 TCTACAGATCAAGAGCACTGTGACGTGGGAGCGCGCGCTAGAGATGATTAAAGC 2274

Qy 2438 TTCCCGGGGCTTAGCACCGGTGAATTTC 2463
Db 2275 TTCCCGGGGCTTAGCACCGGTGAATTTC 2300
RESULT 13
AR373388
LOCUS AR373388 2300 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 83 from patent US 6602705.
ACCESSION AR373388
VERSION AR373388.1 GI:40075491
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2300)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source location/Qualifiers
1..2300
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 82.0%; Score 2019.2; DB 6; Length 2300;
Best Local Similarity 93.2%; Pred. No. 2e-164;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
Qy 170 GCGGCAAGAGGCGCAACGATGAAGACTGACCGAGCGCAAGCCAACTTTCGCG 229
Db 1 GCGGCGCGAGAGACCAATGAAGATTGACTGAGAGAGAGGCTAATTTCTTCGCG 60
Qy 230 AGGACTGGCTTTCCTCCGAGGCAAGGCCCGGAGTTCCCGAGGAGCAAGACCGGCGCA 289
Db 61 AGGACTGGCTTTCCTGAGGCGCAAGGCCCGGAGTTGAGAGCGAGAGACCGGCGCA 120
Qy 290 ACAGCCCCACAGCGCGGAGTGCAGGTCGCGCG-----ACAACCCCGCAGGAG 343
Db 121 ACAGCCCCACAGCGCGCGAGCTGCAAGTGTGGGGCGCGAGAACCAACCTGAGCGAG 180
Qy 344 CCGGCGCGAGCGCAAGGCGACCTTG-----AATCTCCCAAGATCACTGTGCAAC 397
Db 181 CCGGCGCGAGCGCGCAAGGCGACCTTGCACTTCACTTCCCAAGATCACTGTGCAAC 240
Qy 398 GCCCTCTGTGATCAAGTGTGGCGGCAATCAAGAGGCGCTGTGACACCGGCG 457
Db 241 GCCCTCTGTGATCAATCAAGATCGGCGGCACTCAAGAGGCGCTGTGACACCGGCG 300
Qy 458 CCGAGGACACCGTGTGAGAGATGAGCTGCCCGCGCAAGTGAAGCCCAAGATGAC 517
Db 301 CCGAGGACACCGTGTGAGAGATGAGATCTGCCCGCAAGTGAAGCCCAAGATGAC 360
Qy 518 GCGGCAATCGGCGCTTCAATCAAGTGTGCGCAGTACGACGATCTGATGAGATGCG 577
Db 361 GCGGCAATCGGCGCTTCAATCAAGTGTGCGCAGTACGACGATCTGATGAGATGCG 420
Qy 578 GCAAGAGGCGCATTCGCGACCGTGTGATGCGGCCCAACCCCGTGAACATCATGCGCGCA 637
Db 421 GCGACAGGCGCATTCGCGACCGTGTGATGCGGCCCAACCCCGTGAACATCATGCGCGCA 480
Qy 638 ACATGCTGACCGAGCTGGGCTGCAACCTTGAACCTTCCCATCAAGCCCATGAGACGCTG 697
Db 481 ACCTGCTGACCGAGATCGGCTGCAACCTTGAACCTTCCCATCAAGCCCATGAGACGCTG 540
Qy 698 CCGTGAAGCTGAAGCCCGCATGAGACGCGCCCAAGGTGAAGAGTGGCTTGAACGAGG 757
Db 541 CCGTGAAGCTGAAGCCCGGAGATGAGACGCGCCCAAGGTGAAGAGTGGCTTGAACGAGG 600
Qy 758 AGAAGATCAAGGCCCTGACCGCCATCTGCGAGAGAGTGAAGAAAGAGGCGCAAGATCAACA 817

Db 2961 TGGCCGAGGCCATGAGCAAGTGAACAACCAATCATATGATGACGCGCAGCAACTGCA 3020
QY 74 AGGCCCCCAAGCGCATCATCAATGCTTCAACTGCGCGCAAGAGGGCCACATCGCCCGCA 133
Db 3021 AGGGCCCCCAAGCGCACATCAATGCTTCAACTGCGCGCAAGAGGGCCACCTGGCCCGCA 3080
QY 134 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGCCACCAAGTGA 193
Db 3081 ACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAGTGGCGCAAGAGGGCCACCAAGTGA 3140
QY 194 AGGACTGACCGAGCGCGCAGGCCAACTTTCGCGCAGAGGACCTGCGCTTCCCCAGGGCA 253
Db 3141 AGGACTGACCGAGCGCGCAGGCCAACTTTCGCGCAGAGGACCTTCCCAAGAGGA 3194
QY 254 AGGCGCGCAGTTCCTCCAGCAGCAGAACCGCGCCAAAGCCCCCAAGCGCGCAGCTGC 313
Db 3195 AGGCGCGCAGTTCCTTCAGAGCAGACAGAGGCCAAAGCCCCCAAGAGAGCTTC 3254
QY 314 AGGTGCGCGG-----CGAACACCCCGCAGAGGGCCGCGCCGAGCGCCAGGGCA--- 364
Db 3255 AGGTTCGGGGAAGAGCAACAACTCCCTCAGAAACAGAGGCCGATGAGCAAGGAATCG 3314
QY 365 ----CCCTGAATTCCTCCCGATCACTGTGGCAGGCGCCCGCTGTGAGCATCAAGTGC 421
Db 3315 TATCTTTAGCTTCCTCAGATCATCTTTGGAGCGACCCCTCGTCACAATTAAGATAG 3374
QY 422 GCGCGCAGATCAAGAGGCGCTGTGAGCAGCGCGCGCAGCAGCACTGTCTGAGAGGA 481
Db 3375 GGGGGCAGTGAAGAGGGCCCTTTCTAGACACCGGGCCGACCGAGCTGAGAGGA 3434
QY 482 TGAAGCTGCGCGCAGATGGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATAGG 541
Db 3435 TGAACCTGCGCGCGCGCTGGAAGCCCAAGATGATCGCGCGCATCGCGGCTTCATAGG 3494
QY 542 TGGCGCAGATACGACCGATCTGATGAGATCTGCGCGCAAGAGGCGCATCGGCACTGTC 601
Db 3495 TGGCGCAGATACGACCGATCTGATGAGATCTGCGCGCAAGAGGCGCATCGGCACTGTC 3554
QY 602 TGAATCGGCGCGCACCCCGTGAACATCATCGCGCGCAACATGTGACCCAGCTGGCTGCA 661
Db 3555 TGGTGGGGCCCAACCCCGTGAACATCATGGCGCGCACTGTGACCCAGATGGCTGCA 3614
QY 662 CCCTGAATTCCTCCATCAGCCCCATGAGACCGTGCCCGTGAAGCTGAAGCCCGCATGG 721
Db 3615 CCCTGAATTCCTCCATCAGCCCCATGAGACCGTGCCCGTGAAGCTGAAGCCCGCATGG 3674
QY 722 AGGCGCGCGCAAGTGAAGCGTGGCCCTGACCGAGAGGAAGATCAAGGCGCTGACCGCA 781
Db 3675 AGGCGCGCGCAAGTGAAGCGTGGCCCTGACCGAGAGGAAGATCAAGGCGCTGTGAGA 3734
QY 782 TCTGCGAGGATGAGAGAGGAGGCAAGTCAACCAAGTCGCGCCCGAGAACCTCTAGA 841
Db 3735 TCTGCGACCGAGATGAGAGAGGAGGCAAGTCAAGCAAGATCGGCGCCCGAGAACCTCTAGA 3794
QY 842 ACAACCCCGTGTTCGCTATCAAGAGAGAGCAGCAACAGTGGCGCAAGCTGTGACT 901
Db 3795 ACAACCCCGTGTTCGCTATCAAGAGAGAGCAGCAACAGTGGCGCAAGCTGTGACT 3854
QY 902 TCGCGAGGCTGAACAGCGCACCGAAGCTTCTGGAGAGTGGCGCTGGGAGATGCCCGACC 961
Db 3855 TCGCGAGGCTGAACAGCGCACCGAAGCTTCTGGAGAGTGGCGCTGGGAGATGCCCGACC 3914
QY 962 CCGCGCGCGCTGAAGAGAGAGAGCGTGAACGCTGTGAGCTGTGGCGACGCTACTTCA 1021
Db 3915 CCGCGCGCGCTGAAGAGAGAGAGCGTGAACGCTGTGAGCTGTGGCGACGCTACTTCA 3974
QY 1022 GCGTGCCTCTGAGCAGAGACTTCGCGAAGTACACCGCTTCACCATCCCGAGCATCAAC 1081
Db 3975 GCGTGCCTCTGAGCAGAGACTTCGCGAAGTACACCGCTTCACCATCCCGAGCATCAAC 4034
QY 1082 AGGAGACCCCGGATTCGCTACAGTACAAAGTGTGCGCCAGGGCTGGAGGGGAGCC 1141
Db 4035 AGGAGACCCCGGATTCGCTACAGTACAAAGTGTGCGCCAGGGCTGGAGGGGAGCC 4094

QY 1142 CCAGATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCAACCCCG 1201
Db 4095 CCGCATCTTCCAGAGCAGCATGACCAAGATCTTGGAGCCCTTTCGCGCGCGCAACCCCG 4154
QY 1202 AGATGATATCTACCA-----GGCCCCCTGTACGTGGGCGAGCACTGAGATCGGC 1255
Db 4155 AGATGATATCTACCAAGATGACATGAGCAACTGTACGTGGGCGAGCACTGAGATCGGC 4214
QY 1256 AGCACCGCGCCAGATGAGAGCTGCGCAGACCTGCTGCGCTTCCACCAACC 1315
Db 4215 AGCACCGCACCAAGATGAGAGAGCTGCGCAGACCTGCTGCGCTTCCACCAACC 4274
QY 1316 CCGAAGAGAGCAG 1375
Db 4275 CCGAAGAGAGCAG 4334
QY 1376 ACAAGTGAACGCTGAGCGCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1435
Db 4335 ACAAGTGAACGCTGAGCGCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 4394
QY 1436 TCCAGAGCTGTGGGCAAGCTGAACCTGCGCGCAGCAGATCTTACCCCGCATCAAGTGC 1495
Db 4395 TCCAGAGCTGTGGGCAAGCTGAACCTGCGCGCAGCAGATCTTACCCCGCATCAAGTGC 4454
QY 1496 GCGAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1555
Db 4455 GCGAGCTGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4514
QY 1555 AGGAGCGGAGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1615
Db 4515 AGGAGCGGAGTGAAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4574
QY 1616 TGTACTACAGCCCGCAGCAAGAGCCTGTGCGCGAGATCCAGAGAGAGAGAGAGAGAG 1675
Db 4575 TGTACTACAGCCCGCAGCAAGAGCCTGTGCGCGAGATCCAGAGAGAGAGAGAGAGAG 4634
QY 1676 GGAAGCTACAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1735
Db 4635 GGAAGCTACAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4694
QY 1736 TGGCGACGCGCGCACCAAG 1795
Db 4695 TGGCGAGCGCGCACCAAG 4754
QY 1796 TGAAGAGCATGATCTTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1855
Db 4755 TGAAGAGCATGATCTTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4814
QY 1856 CCTGGAGAGCCTGTGAGCAGACTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1915
Db 4815 CCTGGAGAGCCTGTGAGCAGACTTACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4874
QY 1916 TGAACACCCCGCTGTGAG 1975
Db 4875 TGAACACCCCGCTGTGAG 4934
QY 1976 CCGAGAGCTTCTACGAG 2035
Db 4935 CCGAGAGCTTCTACGAG 4994
QY 2036 ACGTGACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2095
Db 4995 ACGTGACCGAGCGGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5054
QY 2096 CCGAGCTGAGAGCAG 2155
Db 5055 CCGAGCTGAGAGCAG 5114
QY 2156 CCGAGCAGCAGATGAGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2215
Db 5115 CCGAGCAGCAGATGAGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5174

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1296.67 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463
Sequence: 1 gtcagccaccacatgcccga.....gggtcagcaccgtgaattc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N_Geneseq_21.*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2463	100.0	2463	6	ABL39960	ABL39960 Synthetic
2	2463	100.0	2463	12	ADM73765	ADM73765 HIV-1 pol
3	2455.4	99.7	2457	8	ACA03547	ACA03547 Synthetic
4	2455.4	99.7	2457	10	ADCI3265	ADCI3265 DNA of HI
5	2442.2	99.2	2469	6	ABL39959	ABL39959 Synthetic
6	2442.2	99.2	2469	12	ADM73764	ADM73764 HIV-1 pol
7	2436.2	98.9	2457	6	ABL39961	ABL39961 Synthetic
8	2436.2	98.9	2457	12	ADM73766	ADM73766 HIV-1 pol
9	2430.2	98.7	2457	8	ACA03548	ACA03548 Synthetic
10	2430.2	98.7	2457	10	ADCI3266	ADCI3266 DNA of HI
11	2422.6	98.4	2445	8	ACA03546	ACA03546 Synthetic
12	2422.6	98.4	2445	10	ADCI3264	ADCI3264 DNA of HI
13	2415.6	98.1	3930	10	ADCI3230	ADCI3230 DNA of HI
14	2414	98.0	3930	10	ADCI3231	ADCI3231 DNA of HI
15	2414	98.0	3930	10	ADCI3232	ADCI3232 DNA of HI
16	2414	98.0	5184	8	ACA03591	ACA03591 Synthetic
17	2414	98.0	5184	10	ADCI3279	ADCI3279 DNA of HI
18	2383.6	96.8	3531	10	ADCI3234	ADCI3234 DNA of HI
19	2382	96.7	3537	10	ADCI3236	ADCI3236 DNA of HI

20	2381	96.7	5145	8	ACA03521	ACA03521 Synthetic
21	2381	96.7	5145	10	ADCI3233	ADCI3233 DNA of HI
22	2371	96.3	3538	10	ADCI3235	ADCI3235 DNA of HI
23	2322.6	94.3	3607	8	ACA03551	ACA03551 Synthetic
24	2322.6	94.3	3607	10	ADCI3269	ADCI3269 DNA of HI
25	2322.6	94.3	3624	8	ACA03550	ACA03550 Synthetic
26	2322.6	94.3	3624	10	ADCI3268	ADCI3268 DNA of HI
27	2304.4	93.6	3597	8	ACA03549	ACA03549 Synthetic
28	2304.4	93.6	3597	10	ADCI3267	ADCI3267 DNA of HI
29	2142	87.0	2466	8	ACA03542	ACA03542 Synthetic
30	2142	87.0	2466	8	ACC78506	ACC78506 HIV p2Pol
31	2122.8	86.2	2472	8	ACA03543	ACA03543 Synthetic
32	2122.8	86.2	2472	8	ACC78507	ACC78507 HIV p2Pol
33	2115.2	85.9	2460	8	ACA03541	ACA03541 Synthetic
34	2115.2	85.9	2460	8	ACC78505	ACC78505 HIV p2Pol
35	2114.2	85.8	3564	8	ACC78488	ACC78488 HIV GagPo
36	2114.2	85.8	3564	8	ACC78489	ACC78489 HIV GagPo
37	2113.6	85.8	4716	8	ACA03522	ACA03522 Synthetic
38	2113.6	85.8	4716	10	ADCI3238	ADCI3238 DNA of HI
39	2109.8	85.7	3999	8	ACC78484	ACC78484 HIV GagCo
40	2108.2	85.6	3999	8	ACC78485	ACC78485 HIV GagCo
41	2108.2	85.6	3999	8	ACC78486	ACC78486 HIV GagCo
42	2108.2	85.6	5283	8	ACA03584	ACA03584 Synthetic
43	2108.2	85.6	5283	8	ACC78529	ACC78529 HIV TatRe
44	2108	85.6	4713	8	ACA03592	ACA03592 Synthetic
45	2108	85.6	4713	10	ADCI3280	ADCI3280 DNA of HI

ALIGNMENTS

RESULT 1
ABL39960
ID ABL39960 standard; DNA; 2463 BP.
XX
XX ABL39960;
XX
XX 15-MAY-2002 (first entry)
XX
XX Synthetic construct PR975YM SEQ ID NO:31.
XX
XX Human immunodeficiency virus type C; antigenic HIV type C protein;
XX immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
XX immunostimulant; gene therapy; gene; ds.
XX
XX Human immunodeficiency virus; type C.
OS
XX Synthetic.
XX
XX WO200204493-A2.
XX
XX 17-JAN-2002.
XX
XX 05-JUL-2001; 2001WO-US021241.
XX
XX 05-JUL-2000; 2000US-00610313.
XX
XX (CHIR) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
XX Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
PI WPI; 2002-154920/20.
XX
XX New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
XX applications including DNA immunization or generation of packaging cell
XX lines, particularly in gene therapy.
XX
XX Claim 1; Fig 9; 23pp; English.
XX
XX The present invention describes expression cassettes comprising a
XX polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
XX type C polypeptides. The expression cassettes comprise any of the HIV
XX type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef

CC (1). (1) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunization, generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL3942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX

Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;

Query Match 100.0%; Score 2463; DB 6; Length 2463;

Best Local Similarity 100.0%; Pred. No. 4.4e-297;

Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGACGCGCCATGAGCGCGCATGAGCGCGCACCGACGCGCAACCTCTGATGCG 60
DB 1 GTGACGCGCCATGAGCGCGCATGAGCGCGCACCGACGCGCAACCTCTGATGCG 60
QY 61 CGACGCACTTTCAGAGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCG 120
DB 61 CGACGCACTTTCAGAGGCGCCCAAGCGCATCATCAAGTCTTCACTGCGGCAAGAGGCG 120
QY 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCAAGAGGCGTCTGGAAGTGCAGAGAG 180
DB 121 CACATGCGCGCGCACTGCGCGCGCGCGCGCGCAAGAGGCGTCTGGAAGTGCAGAGAG 180
QY 181 GCGCGACGAGTGAAGGACTGCAACGAGGCGCGCAACCTTCTTCGCGAGGACCTGCGC 240
DB 181 GCGCGACGAGTGAAGGACTGCAACGAGGCGCGCAACCTTCTTCGCGAGGACCTGCGC 240
QY 241 TTCCCGCAGAGGCAAGGCGCGCGAGTTCCCGCAGAGCAGAGACCGCGCCCAAGCGCCAC 300
DB 241 TTCCCGCAGAGGCAAGGCGCGCGAGTTCCCGCAGAGCAGAGACCGCGCCCAAGCGCCAC 300
QY 301 AGCGCGGAGCTGCAAGTGCAGGCGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGAG 360
DB 301 AGCGCGGAGCTGCAAGTGCAGGCGCGCAACCGCGCGAGGCGCGCGCGCGCGCGCGAG 360
QY 361 GCGACCGCTGAACTTCCCGCAGATCAACCTGTCGACAGCGCGCGCTGTGAGCATCAAGTG 420
DB 361 GCGACCGCTGAACTTCCCGCAGATCAACCTGTCGACAGCGCGCGCTGTGAGCATCAAGTG 420
QY 421 GCGCGCGAGATCAAGAGGCGCTGTGTCACACCGCGCGCGCAACCGCTGTGAGAGAG 480
DB 421 GCGCGCGAGATCAAGAGGCGCTGTGTCACACCGCGCGCGCAACCGCTGTGAGAGAG 480
QY 481 ATGAGCGCTGCGCGCAAGTGAAGCGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
DB 481 ATGAGCGCTGCGCGCAAGTGAAGCGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
QY 541 GTGCGCGAGTACGACGAGATCTGATTCGAGATCTGCGCGCAAGAGCGCATTCGCGACCGTG 600
DB 541 GTGCGCGAGTACGACGAGATCTGATTCGAGATCTGCGCGCAAGAGCGCATTCGCGACCGTG 600
QY 601 CTGATCGCGCGCACCGCGGTGAACATCATCGCGCGCGCAACATCTGACCGAGCTGAGCTGC 660
DB 601 CTGATCGCGCGCACCGCGGTGAACATCATCGCGCGCGCAACATCTGACCGAGCTGAGCTGC 660
QY 661 ACCCTGAATCTTCCCATCATGAGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCATG 720
DB 661 ACCCTGAATCTTCCCATCATGAGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCATG 720
QY 721 GACGCGCGCGCAAGTGAAGAGTGCCTTCGACCGAGAGAAATGAAAGGCGCTTGAACCGCG 780
DB 721 GACGCGCGCGCAAGTGAAGAGTGCCTTCGACCGAGAGAAATGAAAGGCGCTTGAACCGCG 780
QY 781 ATCTCGAGAGAGATGAGAGAGGCGCAAGATCAACAGATCGCGCGCGCGAGAACCGCTTAC 840
DB 781 ATCTCGAGAGAGATGAGAGAGGCGCAAGATCAACAGATCGCGCGCGCGAGAACCGCTTAC 840
QY 841 AACACCGCGCTGTGCGCATCAAGAAAGAGACAGCAACCAAGTGCAGAGCTGTGTGAC 900
DB 841 AACACCGCGCTGTGCGCATCAAGAAAGAGACAGCAACCAAGTGCAGAGCTGTGTGAC 900

QY 901 TTCCGCGAGCTGAACAAGCGACCCAGGACTTCTTGGAGGTGACAGCTGCGCATCCCCAC 960
DB 901 TTCCGCGAGCTGAACAAGCGACCCAGGACTTCTTGGAGGTGACAGCTGCGCATCCCCAC 960
QY 961 CCCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTCTGAAGTGGGCGACCGCTTACTTC 1020
DB 961 CCCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTCTGAAGTGGGCGACCGCTTACTTC 1020
QY 1021 AGCGTCCCTTGAAGAGAGATTCCTCGCAAGTACACCGCGCTTCAACATCCCGAGCATCAAC 1080
DB 1021 AGCGTCCCTTGAAGAGAGATTCCTCGCAAGTACACCGCGCTTCAACATCCCGAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCGATCCGCTTACAGATCAACAGTGCCTGCGCCAGAGGCTGGAAGGCGAC 1140
DB 1081 AACGAGACCCCGGCGATCCGCTTACAGATCAACAGTGCCTGCGCGCCAGAGGCGAC 1140
QY 1141 CCAGCATCTTTCAGAGAGAGATGAACCAAGTCTGAGAGCGCTTCCGCGCGCGCAACCGC 1200
DB 1141 CCAGCATCTTTCAGAGAGAGATGAACCAAGTCTGAGAGCGCTTCCGCGCGCGCAACCGC 1200
QY 1201 GAGATCGTGAATCTACAGAGCGCGCGCTGTACGTCGCGAGCGACCTGAGAGATCGCGCAGCAC 1260
DB 1201 GAGATCGTGAATCTACAGAGCGCGCGCTGTACGTCGCGAGCGACCTGAGAGATCGCGCAGCAC 1260
QY 1261 CGCGCGAGATTCGAGAGAGCTGCGCAAGACCTGCTGCGCTGGGCGCTTCAACACCGCGAC 1320
DB 1261 CGCGCGAGATTCGAGAGAGCTGCGCAAGACCTGCTGCGCTGGGCGCTTCAACACCGCGAC 1320
QY 1321 AAGAGACACAGAGAGAGCGCGCGCTTCTGTGATGAGGCTACAGAGCTGACCCCGACAG 1380
DB 1321 AAGAGACACAGAGAGAGCGCGCGCTTCTGTGATGAGGCTACAGAGCTGACCCCGACAG 1380
QY 1381 TGGACCTGTGAGCGCATGAGCTGCGCGAGAGAGAGACTGGAACCTGTGAAGCATTCAG 1440
DB 1381 TGGACCTGTGAGCGCATGAGCTGCGCGAGAGAGAGACTGGAACCTGTGAAGCATTCAG 1440
QY 1441 AAGCTGTGAGGAGAGCTGAACCTGGGCGAGCGCATCTACCGCGGATCAAGGTGCGCGAG 1500
DB 1441 AAGCTGTGAGGAGAGCTGAACCTGGGCGAGCGCATCTACCGCGGATCAAGGTGCGCGAG 1500
QY 1501 CTGTGCAAGCTGCTGCGCGCGCGCAAGGCGCTTGAACCATCTGTGCGCGCTTGAACCGAGAG 1560
DB 1501 CTGTGCAAGCTGCTGCGCGCGCGCAAGGCGCTTGAACCATCTGTGCGCGCTTGAACCGAGAG 1560
QY 1561 GCGGAGCTGAGAGCTGCGCGAGAACCGCGAGATCTTGCAGAGCGCTGTGATC 1620
DB 1561 GCGGAGCTGAGAGCTGCGCGAGAACCGCGAGATCTTGCAGAGCGCTGTGATC 1620
QY 1621 TACGACCGCGAGAGAGCTGTGCGAGATTCAGAAAGCAGAGCGACGACAGATGAGACC 1680
DB 1621 TACGACCGCGAGAGAGCTGTGCGAGATTCAGAAAGCAGAGCGACGACAGATGAGACC 1680
QY 1681 TACGAGATCTTACAGAGAGCGCTTCAAGAACCTGAAGACCGCGCATGACCGCAAGATGCGC 1740
DB 1681 TACGAGATCTTACAGAGAGCGCTTCAAGAACCTGAAGACCGCGCATGACCGCAAGATGCGC 1740
QY 1741 ACCGCGCAACCAACGAGCTGAAGAGCTGACCGAGCGCGTGCAGAAATGCGCATGAGAG 1800
DB 1741 ACCGCGCAACCAACGAGCTGAAGAGCTGACCGAGCGCGTGCAGAAATGCGCATGAGAG 1800
QY 1801 AGCATCGGATCTGCGGCGAGAGCGCGCGCAAGTTCGCGCTGCGCATCAAGAGAGAGCTGAG 1860
DB 1801 AGCATCGGATCTGCGGCGAGAGCGCGCGCGCAAGTTCGCGCTGCGCATCAAGAGAGAGCTGAG 1860
QY 1861 GAGAGCTGTGAGACGACTGAGAGGCGACCTTGAATCCCGAGTGGAGATTCTGTGAAC 1920
DB 1861 GAGAGCTGTGAGACGACTGAGAGGCGACCTTGAATCCCGAGTGGAGATTCTGTGAAC 1920
QY 1921 ACCCGCGCGCTGTGAGAGCTGTGTGTAACAGCTGAGAGAGAGCGCATCATCGCGCGAG 1980
DB 1921 ACCCGCGCGCTGTGAGAGCTGTGTGTAACAGCTGAGAGAGAGCGCATCATCGCGCGAG 1980

```
QY 1981 ACCCTTCAAGTGAAGCGCGCCGCAACCGCGAGACCAAGATCGGGAAGCGCGCTAACGCG 2040
DB 1981 ACCCTTCAAGTGAAGCGCGCCGCAACCGCGAGACCAAGATCGGGAAGCGCGCTAACGCG 2040
QY 2041 ACCGACCGGGGCGCGGCAAGATCGTGAACCTGACCCGAGACCAACCAAGAGACCGAG 2100
DB 2041 ACCGACCGGGGCGCGGCAAGATCGTGAACCTGACCCGAGACCAACCAAGAGACCGAG 2100
QY 2101 CTGCAAGCGCATCAAGCTGCGCCCTGCAAGACAGCGGCAAGAGTGAACCTGTCAGCCAG 2160
DB 2101 CTGCAAGCGCATCAAGCTGCGCCCTGCAAGACAGCGGCAAGAGTGAACCTGTCAGCCAG 2160
QY 2161 ACCGAGTAAAGCCCTGCGGATCATCCAGGCCCGACCCGCAAGAGCGAGAGCGAGCTGCG 2220
DB 2161 ACCGAGTAAAGCCCTGCGGATCATCCAGGCCCGACCCGCAAGAGCGAGAGCGAGCTGCG 2220
QY 2221 AACCGATCATGAGAGCTGATCAAGAGAGAGAGTGAACCTGAGCTGCGGCGCCGCG 2280
DB 2221 AACCGATCATGAGAGCTGATCAAGAGAGAGAGTGAACCTGAGCTGCGGCGCCGCG 2280
QY 2281 CACAAAGGCGATCGCGCGCAACGACGATCGAACAGCTGTCAGCAAGGCGATTCGCAAG 2340
DB 2281 CACAAAGGCGATCGCGCGCAACGACGATCGAACAGCTGTCAGCAAGGCGATTCGCAAG 2340
QY 2341 GTGCTGTCTTGAAGCGGATCGATGCGCGGATCTGATCTAACAGTACATGACACGCTG 2400
DB 2341 GTGCTGTCTTGAAGCGGATCGATGCGCGGATCTGATCTAACAGTACATGACACGCTG 2400
QY 2401 TACGTGGGAGAGCGCGCGCTAGGATCGATTAAAGCTTCCGCGGCGCTAGCACCGGTGA 2460
DB 2401 TACGTGGGAGAGCGCGCGCTAGGATCGATTAAAGCTTCCGCGGCGCTAGCACCGGTGA 2460
QY 2461 TTC 2463
DB 2461 TTC 2463
```

RESULT 2

ADM73765
ID ADM73765 standard; DNA; 2463 BP.

AC ADM73765;

DT 03-JUN-2004 (first entry)

DE HIV-1 polynucleotide #8.

KM HIV-1; gene; de; HIV pol; immune response; DNA immunisation;

KW HIV type C protein; immunostimulant.

OS Human immunodeficiency virus 1.

PN US2003223961-A1.

PD 04-DEC-2003;

PF 05-JUL-2001; 2001US-00899575.

PR 05-JUL-2000; 2000US-00610313.

PA (MEGE/) MEGEDE J Z.

PA (BARN/) BARNETT S W.

PA (ENG/) ENGELBRECHT S.

PA (RENS/) RENSBURG E J V.

PI Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;

DR WPI; 2004-060515/06.

XX New expression cassette comprising a polynucleotide sequence encoding an
PT HIV pol polypeptide, useful in eliciting an immune response, in DNA
PT immunization, generating of packaging cell lines or in producing HIV type
PT C proteins.

```
XX  
PS Claim 1; SEQ ID NO 31; 160pp; English.  
XX  
CC The invention relates to an expression cassette comprising a  
CC polynucleotide sequence encoding an HIV Pol polypeptide. The invention  
CC also relates to a recombinant expression system for use in a host cell  
CC comprising an expression cassette, where the polynucleotide sequence  
CC further comprises control elements capable of driving expression in the  
CC selected host cell, a cell comprising an expression cassette where the  
CC polynucleotide sequence further comprises control elements compatible  
CC with the expression in the cell and a composition for generating an  
CC immunological response, comprising an expression cassette. The expression  
CC cassette and the methods of the invention are useful in eliciting an  
CC immune response, in DNA immunisation, in generation of packaging cell  
CC lines and in producing HIV type C proteins. This sequence represents an  
CC HIV-1 polynucleotide of the invention.  
SQ Sequence 2463 BP; 567 A; 835 C; 759 G; 302 T; 0 U; 0 Other;
```

Query Match 100.0%; Score 2463; DB 12; Length 2463;

Best Local Similarity 100.0%; Pred. No. 4.4e-297;

Matches 2463; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGACGCGCACCATGCGCGAGGCGCATGAGCCAGGCGCACAGGCGCAATCCGATGACG 60
DB 1 GTGACGCGCACCATGCGCGAGGCGCATGAGCCAGGCGCACAGGCGCAATCCGATGACG 60
QY 61 CGCAGCAACTTCAGAGGCGCCCGCAAGCGCATCATAGTCTTCACTGCGCGCAAGAGGCG 120
DB 61 CGCAGCAACTTCAGAGGCGCCCGCAAGCGCATCATAGTCTTCACTGCGCGCAAGAGGCG 120
QY 121 CATATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGCTGCTGAAAGTGGCGCAAGAG 180
DB 121 CATATGCGCCGCACTGCGCGCGCCCGCGCAAGAGGCTGCTGAAAGTGGCGCAAGAG 180
QY 181 GGCACACGAGTAAAGATGACGACCGAGCGCGCAACTTCTTCGCGAGGACCTGCGCC 240
DB 181 GGCACACGAGTAAAGATGACGACCGAGCGCGCAACTTCTTCGCGAGGACCTGCGCC 240
QY 241 TTCGCCCGAGGCGAGGCGCGCGAGTTCGCCAGAGCAGACCGCGCAACGCGCCACCC 300
DB 241 TTCGCCCGAGGCGAGGCGCGCGAGTTCGCCAGAGCAGACCGCGCAACGCGCCACCC 300
QY 301 AGCCGCGAGCTGCAAGTGCAGCGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCG 360
DB 301 AGCCGCGAGCTGCAAGTGCAGCGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCG 360
QY 361 GGCACCTGGAATTTCCCGCGAGTCAACCTCTGTGGAGCGCGCCCTGTGTGAGCATCAAG 420
DB 361 GGCACCTGGAATTTCCCGCGAGTCAACCTCTGTGGAGCGCGCCCTGTGTGAGCATCAAG 420
QY 421 GCGCGCGAGATCAAGAGGCGCGTGTGACACCGCGCGCGAGACACCGTGTGAGGAG 480
DB 421 GCGCGCGAGATCAAGAGGCGCGTGTGACACCGCGCGCGAGACACCGTGTGAGGAG 480
QY 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATGCGCGCATGCGCGCTTCATCAAG 540
DB 481 ATGAGCTGCGCGCGCAAGTGAAGCCCAAGATGATGCGCGCATGCGCGCTTCATCAAG 540
QY 541 GTGCGCGAGTAAAGACCAAGTCTGTATCGAGATCTGCGGCAAGAGCGCATTCGCGCA 600
DB 541 GTGCGCGAGTAAAGACCAAGTCTGTATCGAGATCTGCGGCAAGAGCGCATTCGCGCA 600
QY 601 CTGATCGCGCGCACCGCGTGAACATCATCGGCGCGCAATCATCGGCTGAGCCAGCTGCG 660
DB 601 CTGATCGCGCGCACCGCGTGAACATCATCGGCGCGCAATCATCGGCTGAGCCAGCTGCG 660
QY 661 ACCCTGAATTTCCCATAGCCCGCATGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 720
DB 661 ACCCTGAATTTCCCATAGCCCGCATGAGACCGTGCCTGTGAAGCTGAAGCCCGCATG 720
QY 721 GACGCGCGCAAGTGAAGAGTGGCGCGTGAACCGAGAGAAATCAAGGCGCTGACCGCG 780
DB 721 GACGCGCGCAAGTGAAGAGTGGCGCGTGAACCGAGAGAAATCAAGGCGCTGACCGCG 780
```

```

Db      721 GACGGCCCCAGGTGAAGCATGTGCCCCCTTGAACCGAGAGAAAGATCAAGGCCCTGACCGCC 780
Qy      781 ATCTGCGAGAGATGAGAGAGAGAGGAGGCAAGATCACCAAGATGAGGCCCGGAGAACCCCTTAC 840
Db      781 ATCTGCGAGAGATGAGAGAGAGAGGAGGCAAGATCACCAAGATGAGGCCCGGAGAACCCCTTAC 840
Qy      841 AACACCCCCGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGGCGGCAAGCTGTGTGAGC 900
Db      841 AACACCCCCGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGGCGGCAAGCTGTGTGAGC 900
Qy      901 TTCCCGGAGCTGAACAAGGCAACCCAGACCTTCTGAGAGGTGAGTGGAGCTTCCCGAC 960
Db      901 TTCCCGGAGCTGAACAAGGCAACCCAGACCTTCTGAGAGGTGAGTGGAGCTTCCCGAC 960
Qy      961 CCCGCGGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGAGCGGCTTACTTC 1020
Db      961 CCCGCGGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGAGCGGCTTACTTC 1020
Qy      1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGCCTTGCACATCCCGACATCAAC 1080
Db      1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGCCTTGCACATCCCGACATCAAC 1080
Qy      1081 AACGAGACCCCGGCGCATCCGCTACCAAGTACAAAGTGTGCCCCAGAGGCTGAGAGGCGAGC 1140
Db      1081 AACGAGACCCCGGCGCATCCGCTACCAAGTACAAAGTGTGCCCCAGAGGCTGAGAGGCGAGC 1140
Qy      1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGAGAGCCCTTCCGCGCGCGCAACCC 1200
Db      1141 CCCAGCATCTTCAAGAGAGAGATGACCAAGATCTGAGAGCCCTTCCGCGCGCGCAACCC 1200
Qy      1201 GAGATGTGATCTTACAGAGCCCGCCCTGTACGTGTGAGAGAGAGAGAGAGTGGAGTCCGAGC 1260
Db      1201 GAGATGTGATCTTACAGAGCCCGCCCTGTACGTGTGAGAGAGAGAGAGAGTGGAGTCCGAGC 1260
Qy      1261 GCGCGCAAGATGAGAGAGAGTGTGCAAGACCTGTGCGGTGTGAGGCTTCAACACCCCGAC 1320
Db      1261 GCGCGCAAGATGAGAGAGAGTGTGCAAGACCTGTGCGGTGTGAGGCTTCAACACCCCGAC 1320
Qy      1321 AAGAGACACAGAGAGAGCCCGCTTCTGTGAGTGGAGTGAAGCTGAGACCCCGACAG 1380
Db      1321 AAGAGACACAGAGAGAGCCCGCTTCTGTGAGTGGAGTGAAGCTGAGACCCCGACAG 1380
Qy      1381 TGGACCGTGAAGCCCATGAGCTGTGCGGAGAGAGAGAGAGTGAACCGTGAACATTCAG 1440
Db      1381 TGGACCGTGAAGCCCATGAGCTGTGCGGAGAGAGAGAGAGTGAACCGTGAACATTCAG 1440
Qy      1441 AAGCTGTGTGGAGAGTGAAGTGTGAGCTGTGCGGAGAGAGAGTGAACCGTGAACATTCAG 1500
Db      1441 AAGCTGTGTGGAGAGTGAAGTGTGAGCTGTGCGGAGAGAGAGTGAACCGTGAACATTCAG 1500
Qy      1501 CTGTGCAAGCTGTGCGCGGAGAGAGAGCCCTGACCGACATGTGTGCGGAGAGAG 1560
Db      1501 CTGTGCAAGCTGTGCGCGGAGAGAGAGCCCTGACCGACATGTGTGCGGAGAGAG 1560
Qy      1561 GCGAGCTGTGAGCTGTGCGGAGAGAGAGAGAGTGTGCGGAGAGAGAGAGAGAGAGAG 1620
Db      1561 GCGAGCTGTGAGCTGTGCGGAGAGAGAGAGAGAGAGTGTGCGGAGAGAGAGAGAGAGAG 1620
Qy      1621 TACGAGACCCAGAGAGAGCTGTGTGCGGAGAGTTCAGAGAGAGAGAGAGAGAGAGAG 1680
Db      1621 TACGAGACCCAGAGAGAGCTGTGTGCGGAGAGTTCAGAGAGAGAGAGAGAGAGAGAG 1680
Qy      1681 TACGAGATCTACAGAGAGCCCTTCAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1740
Db      1681 TACGAGATCTACAGAGAGCCCTTCAAGAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy      1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db      1741 ACCGCGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy      1801 AGCATGTGTATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db      1801 AGCATGTGTATCTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860

```

```

Qy      1861 GAGACTGTGTGAGACCGACCTACTGAGAGGCACTTGGATCCCGAGTGGAGAGTTCGTGAC 1920
Db      1861 GAGACTGTGTGAGACCGACCTACTGAGAGGCACTTGGATCCCGAGTGGAGAGTTCGTGAC 1920
Qy      1921 ACCCCCCCTGTGTAAAGCTGTGTGACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db      1921 ACCCCCCCTGTGTAAAGCTGTGTGACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy      1981 ACCTTCTAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Db      1981 ACCTTCTAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy      2041 ACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Db      2041 ACCGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy      2101 CTGAGAGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Db      2101 CTGAGAGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy      2161 AGCGAGTACGCGCTGTGTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db      2161 AGCGAGTACGCGCTGTGTGAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy      2221 AACGAGATCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Db      2221 AACGAGATCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy      2281 CAAAGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db      2281 CAAAGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy      2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Db      2341 GTGCTGTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy      2401 TACGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Db      2401 TACGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2460
Qy      2461 TTC 2463
Db      2461 TTC 2463

RESULT 3
ACAO3547
ID      ACAA03547 standard; DNA; 2457 BP.
XX      ACAA03547;
AC      ACAA03547;
XX      ACAA03547;
DT      22-MAY-2003 (first entry)
XX      22-MAY-2003 (first entry)
DE      Synthetic DNA encoding immunogenic HIV peptide #30.
XX      Synthetic DNA encoding immunogenic HIV peptide #30.
KW      Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW      gene therapy; packaging cell line; humoral immune response;
KW      cellular immune response; gene delivery vector; DNA immunisation; de.
OS      Synthetic.
XX      Synthetic.
PN      MO2003004657-A1.
PD      16-JAN-2003.
XX      16-JAN-2003.
PF      05-JUL-2002; 2002WO-US021421.
XX      05-JUL-2002; 2002WO-US021421.
PR      31-AUG-2001; 2001US-0303192P.
PR      31-AUG-2001; 2001US-0316860P.
PR      16-JAN-2002; 2002US-0349728P.
PR      16-JAN-2002; 2002US-0349793P.
PR      16-JAN-2002; 2002US-0349871P.

```

XX (CHTR) CHIRON CORP.
 XX Zur Megede J, Barnett SW, Lian Y;
 XX WPI: 2003-221602/21.
 DR New synthetic polynucleotides encoding antigenic HIV type B and/or type C
 XX polypeptides, useful as immunogenic compositions or vaccines for
 PT generating humoral or cellular immune responses against HIV in a subject,
 XX especially humans.
 XX Example 1; Fig 35; 262pp; English.
 XX The invention describes a synthetic polynucleotide encoding 2 or more
 CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
 CC derived from different HIV subtypes. The polynucleotide is useful for
 CC immunisation, generation of packaging cell lines, or production of HIV
 CC polypeptides. The polynucleotide and its encoded proteins are useful as
 CC immunogenic compositions or vaccines for generating humoral or cellular
 CC immune responses against HIV in a subject, or for inducing neutralising
 CC antibodies against HIV. The gene delivery vector comprising the
 CC polynucleotide is also useful for DNA immunisation of, or for generating
 CC an immune response (e.g. a humoral or cellular immune response) in, a
 CC subject such as a mammal, particularly a human. This sequence encodes a
 CC human immunodeficiency virus immunogenic peptide
 XX
 SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;
 Query Match 99.7%; Score 2455.4; DB 8; Length 2457;
 Best Local Similarity 100.0%; Pred. No. 3,9e-296;
 Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 601 CTGATGGGCCCCACCCCGTGAAATCATGCGCGCAACATGCTGACCAAGCTGGGCTGC 660
 DB 601 CTGATGGGCCCCACCCCGTGAAATCATGCGCGCAACATGCTGACCAAGCTGGGCTGC 660
 QY 661 ACCCTGAACCTTCCCATCAGCCCATCGAAGCCGTGCCGTGAACCTGAAGCCCGCATG 720
 DB 661 ACCCTGAACCTTCCCATCAGCCCATCGAAGCCGTGCCGTGAACCTGAAGCCCGCATG 720
 QY 721 GACGGCCCCAAGGTGAAGAGATGAGGCCCCCTGACCCGAGAGAAATATGAAGCCCTGACCCGC 780
 DB 721 GACGGCCCCAAGGTGAAGAGATGAGGCCCCCTGACCCGAGAGAAATATGAAGCCCTGACCCGC 780
 QY 781 ATCTGCGAGAGATGAGAGAGAGGCGAAGATCATCAAGATGAGGCCCGGAGAACCCCTAC 840
 DB 781 ATCTGCGAGAGATGAGAGAGAGGCGAAGATCATCAAGATGAGGCCCGGAGAACCCCTAC 840
 QY 841 AACACCCCGTGTTCGCCATCAAGAGAGAGACAGACCAACATGCGCAAGCTGTGTGAC 900
 DB 841 AACACCCCGTGTTCGCCATCAAGAGAGAGACAGACCAACATGCGCAAGCTGTGTGAC 900
 QY 901 TTCGCGAGCTGAACAAGCGCACCAAGACCTTCTGGAGGTGACAGCTGGGCACTCCCCAC 960
 DB 901 TTCGCGAGCTGAACAAGCGCACCAAGACCTTCTGGAGGTGACAGCTGGGCACTCCCCAC 960
 QY 961 CCGCGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGACCGCTACTTC 1020
 DB 961 CCGCGCGCTGAAGAAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGACCGCTACTTC 1020
 QY 1021 AGCGTGCCCTTGAAGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGACATCAAC 1080
 DB 1021 AGCGTGCCCTTGAAGAGAGCTTCCGCAAGTACACCGCTTACCATCCCGACATCAAC 1080
 QY 1081 AAGAGAGCCCGGCAATCCGCTACAGATCAACAGTGTGCTGCCAAGGGCTGAAGGGCAGC 1140
 DB 1081 AAGAGAGCCCGGCAATCCGCTACAGATCAACAGTGTGCTGCCAAGGGCTGAAGGGCAGC 1140
 QY 1141 CCAGCATCTTCAAGAGAGAGATGAACAGATCTGAGACCCCTTCCGCGCGGCAACCCC 1200
 DB 1141 CCAGCATCTTCAAGAGAGAGATGAACAGATCTGAGACCCCTTCCGCGCGGCAACCCC 1200
 QY 1201 GAGATGATCTACAGAGGCCCCCTGTAGTGTGGGCAAGGACCTGAGATTCGGCGACAC 1260
 DB 1201 GAGATGATCTACAGAGGCCCCCTGTAGTGTGGGCAAGGACCTGAGATTCGGCGACAC 1260
 QY 1261 CGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGTGGGCTTACCAACCCCTGAC 1320
 DB 1261 CGCGCAAGATGAGAGAGCTGCGCAAGACCTGTGCGTGGGCTTACCAACCCCTGAC 1320
 QY 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGTGATGTGGCTTACAGGCTGCAACCCCGACAG 1380
 DB 1321 AAGAGACCAAGAGAGGCCCCCTTCTGTGTGATGTGGCTTACAGGCTGCAACCCCGACAG 1380
 QY 1381 TGGACCGTGCAGCCCATGAGGTGCGCGAAGAGAGAGCTGACCGTGAACGATCTCAG 1440
 DB 1381 TGGACCGTGCAGCCCATGAGGTGCGCGAAGAGAGAGCTGACCGTGAACGATCTCAG 1440
 QY 1441 AAGCTGTGTGGCAAGCTGAATGTGGCGACGACGATCTACCCCGGCAATCAAGGTGCCAG 1500
 DB 1441 AAGCTGTGTGGCAAGCTGAATGTGGCGACGACGATCTACCCCGGCAATCAAGGTGCCAG 1500
 QY 1501 CTGTGCAAGCTGTGCGCGCGCAAGGCCCCCTGACCGACATGTGTGCCCTGACCCGAGAG 1560
 DB 1501 CTGTGCAAGCTGTGCGCGCGCAAGGCCCCCTGACCGACATGTGTGCCCTGACCCGAGAG 1560
 QY 1561 GCGAGAGTGAAGTGGCGAGAAACCGGAGATCTTGCAGGAGCGGTGACCGGCTGTAC 1620
 DB 1561 GCGAGAGTGAAGTGGCGAGAAACCGGAGATCTTGCAGGAGCGGTGACCGGCTGTAC 1620
 QY 1621 TACGACCCAGCAAGAGACTGTGTGCGAGATTCAGAAACAGGCGCAAGCTGTGACCC 1680
 DB 1621 TACGACCCAGCAAGAGACTGTGTGCGAGATTCAGAAACAGGCGCAAGCTGTGACCC 1680
 QY 1681 TACCAATCTACAGAGAGCCCTTCAAGAACTGAAGACCGGCAAGTACCGCAAGATGCCG 1740

```

Db      1681 TACGAGATCTACCAAGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCCAAGATGCGC 1740
Qy      1741 ACCGCCACACCAAGAGAGCTGAAAGACGCTGACCGAGCCCTGACAAAGATCCGCCATGAG 1800
Db      1741 ACCGCCACACCAAGAGAGCTGAAAGACGCTGACCGAGCCCTGACAAAGATCCGCCATGAG 1800
Qy      1801 AGCATGATGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAGAGACTGG 1860
Db      1801 AGCATGATGATCTGGGCGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAGAGACTGG 1860
Qy      1861 GAGACCTGTGAGACCGACTACTGCGAGCGCCACCTGATCCCGAGTGGAGTTCTGTAAC 1920
Db      1861 GAGACCTGTGAGACCGACTACTGCGAGCGCCACCTGATCCCGAGTGGAGTTCTGTAAC 1920
Qy      1921 ACCCCCCCTGCTGTAAGCTGTGTAACCGAGCTGAGAAAGAACCCATCTCCGCCCGAG 1980
Db      1921 ACCCCCCCTGCTGTAAGCTGTGTAACCGAGCTGAGAAAGAACCCATCTCCGCCCGAG 1980
Qy      1981 ACCTTCTAGTGAACGCGCCGCAACCGGAGACCAAGATGCGCAAGGCGGCTAACGTG 2040
Db      1981 ACCTTCTAGTGAACGCGCCGCAACCGGAGACCAAGATGCGCAAGGCGGCTAACGTG 2040
Qy      2041 ACCGACCGGCGCGCGCAAGATCTGAGCTGACCGAGACCAACCAAGAAAGCCGAG 2100
Db      2041 ACCGACCGGCGCGCGCAAGATCTGAGCTGACCGAGACCAACCAAGAAAGCCGAG 2100
Qy      2101 CTGAGAGCCATCTGAGCTGAGCTGAGAGCAAGCGGAGAGGTGAACTCTGACCGAC 2160
Db      2101 CTGAGAGCCATCTGAGCTGAGCTGAGAGCAAGCGGAGAGGTGAACTCTGACCGAC 2160
Qy      2161 AGCCAGTACGCGCTGAGCTGATCTCAAGCGCCAGCCGCAAGAGCGAGAGCGAGCTGTG 2220
Db      2161 AGCCAGTACGCGCTGAGCTGATCTCAAGCGCCAGCCGCAAGAGCGAGAGCGAGCTGTG 2220
Qy      2221 AACGAGTATCTGAGAGCTGATCTCAAGAAAGAAAGGTGTACTGAGCTGGTGCCTG 2280
Db      2221 AACGAGTATCTGAGAGCTGATCTCAAGAAAGAAAGGTGTACTGAGCTGGTGCCTG 2280
Qy      2281 CACAAAGGAGATGCGCGGCAACGAGATCTGAGCTGTGAGCAAGGCGATCTCGCAAG 2340
Db      2281 CACAAAGGAGATGCGCGGCAACGAGATCTGAGCTGTGAGCAAGGCGATCTCGCAAG 2340
Qy      2341 GTGCTGTTCTTGGACGCGATCGATGAGCGGATCTGATCTAACAGATGAGCAAGACTG 2400
Db      2341 GTGCTGTTCTTGGACGCGATCGATGAGCGGATCTGATCTAACAGATGAGCAAGACTG 2400
Qy      2401 TACGTGGGAGCGCGGCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457
Db      2401 TACGTGGGAGCGCGGCTTAGATCGATTAAAGCTTCCCGGGCTAGCACCGGT 2457

```

RESULT 4
ADCL3265
ID ADCL3265 standard; DNA; 2457 BP.

```

AC      ADCL3265;
XX      18-DEC-2003 (first entry)
XX      DNA of HIV construct p2Pol-opt-ym_C seq ID NO 44.
XX      expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX      Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
OS      Human immunodeficiency virus.
XX      WO2003004620-A2.
XX      16-JAN-2003.
XX      05-JUL-2002; 2002WO-US021420.
PF      XX
XX

```

```

PR      05-JUL-2001; 2001US-0303192P.
PR      31-AUG-2001; 2001US-0316860P.
PR      16-JAN-2002; 2002US-0349871P.
XX      (CHIR ) CHIRON CORP.
PA      (UYST-) UNITV STELLENBOSCH.
XX      Zur Megede J, Barnett SM, Llan Y, Engelbrecht S, Van Rensburg RJ,
XX      MPI; 2003-221593/21.
XX      New expression cassette comprising a polynucleotide sequence encoding a
XX      polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX      PT Prot, or Rev polypeptide, useful for immunisation, or generating
XX      packaging cell lines.
XX      Disclosure; Fig 41; 301pp; English.
XX      The invention relates to a novel expression cassette comprising a
XX      polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX      Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX      expression cassette can be used to treat HIV type C by gene therapy or
XX      used in the development of a vaccine. The gene delivery vector is
XX      administered intramuscularly, intravenously, intradermally,
XX      subcutaneously, intradermally, transdermally, intranasally,
XX      intrarectally, orally or intravenously. The expression cassette is useful
XX      for immunisation, generating packaging cell lines and producing HIV
XX      polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX      Type C related sequence of the invention.
SQ      Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;
Query Match          99.7%; Score 2455.4; DB 10; Length 2457;
Blast Local Similarity 100.0%; Pred. No. 3.9e-296;
Matches 2456; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy      1 GTGAGCGCACCATGCGCGAGGCGCATGAGCGGCGCACCGGCGCAACTCTGATGCGAG 60
Db      1 GTGAGCGCACCATGCGCGAGGCGCATGAGCGGCGCACCGGCGCAACTCTGATGCGAG 60
Qy      61 CGGAGCACTTCAAGAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGCG 120
Db      61 CGGAGCACTTCAAGAGGCGCCCAAGCGCATCATCAAGTGTCTCACTGCGGCAAGAGGCG 120
Qy      121 CACATGCGCGCGAACTTGCCTGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGCGGCAAGAG 180
Db      121 CACATGCGCGCGAACTTGCCTGCGCGCGCGCGCGCAAGAGGCTGTGAAAGTGCGGCAAGAG 180
Qy      181 GGCACACAGATGAAGACTGACACGAGCGCGCAAGCTTCTTCCGAGAGGACTGCGC 240
Db      181 GGCACACAGATGAAGACTGACACGAGCGCGCAAGCTTCTTCCGAGAGGACTGCGC 240
Qy      241 TTCCCCCAAGGCGAGGCGCGGAGTTTCCCAAGAGCAAGAACCGCGCCAAAGCCCGACC 300
Db      241 TTCCCCCAAGGCGAGGCGCGGAGTTTCCCAAGAGCAAGAACCGCGCCAAAGCCCGACC 300
Qy      301 AGCGCGAGCTGCAAGTGTGCGCGCGCAACACCCCGCAAGAGCGCGCGCGCGAGCGCGAG 360
Db      301 AGCGCGAGCTGCAAGTGTGCGCGCGCAACACCCCGCAAGAGCGCGCGCGCGAGCGCGAG 360
Qy      361 GGCACCTGGAATCTTCCCGAGATCACTCTGTGCGAGCGCGCGCTGTGAGATCAAGAGTG 420
Db      361 GGCACCTGGAATCTTCCCGAGATCACTCTGTGCGAGCGCGCGCTGTGAGATCAAGAGTG 420
Qy      421 GCGGCGCAAGTCAAGAGGCGCTGTGAGCAACCGCGCGCGAGCAACCTGTCTGAGAGAG 480
Db      421 GCGGCGCAAGTCAAGAGGCGCTGTGAGCAACCGCGCGCGAGCAACCTGTCTGAGAGAG 480
Qy      481 ATGAGCTTGTCCCGGCAAGTGAAGCCCAAGATGATCTGCGCGGATGCGCGGCTTTCATCAG 540
Db      481 ATGAGCTTGTCCCGGCAAGTGAAGCCCAAGATGATCTGCGCGGATGCGCGGCTTTCATCAG 540
Qy      541 GTGCGCGATGACCAAGATCTGATCGAGATCTGCGGCAAGAGCGCATCGGACCGTG 600

```


Db 541 GTGCGCAATACGACCAAGATCTTGATCGAGATCTGGGCAAGAGGCCATCGGACCGTG 600
Qy 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGGCAATGCTGACCCAGCTGGGCTGC 660
Db 601 CTGATCGGCCCCACCCCGGTGAACATCATCGGCGGCAATGCTGACCCAGCTGGGCTGC 660
Qy 661 ACCCTGAATTTTCCCATCAGCCCATCGAGACCTGTGACCTGGTAAGTGAAGCCCGGCA 720
Db 661 ACCCTGAATTTTCCCATCAGCCCATCGAGACCTGTGACCTGGTAAGTGAAGCCCGGCA 720
Qy 721 GACGGCCCCAAGGTGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGGC 780
Db 721 GACGGCCCCAAGGTGAAGAGTGGCCCTGACCGAGAGAGATCAAGGCCCTGACCGGC 780
Qy 781 ATCTGCGAGAGATGAGAGAGAGGCGCAAGATCAACAGATCGGCCCCGAGAACCCCTAC 840
Db 781 ATCTGCGAGAGATGAGAGAGAGGCGCAAGATCAACAGATCGGCCCCGAGAACCCCTAC 840
Qy 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGAGAC 900
Db 841 AACACCCCGGTGTTCCCATCAAGAGAGAGACGACCAAGTGGCGCAAGCTGTGAGAC 900
Qy 901 TTCCGGAAGCTGAACAAAGGCGACCCGAGACTTCTGGAGGTGACGCTGGGCAATCCGCCAC 960
Db 901 TTCCGGAAGCTGAACAAAGGCGACCCGAGACTTCTGGAGGTGACGCTGGGCAATCCGCCAC 960
Qy 961 CCGCGCGGCTGAGAGAGAGAGACGCTGACCGTCTGACCGTGGCGACCGCTACTTTC 1020
Db 961 CCGCGCGGCTGAGAGAGAGAGACGCTGACCGTCTGACCGTGGCGACCGCTACTTTC 1020
Qy 1021 AGCGTCCCCCTGAGCAAGAGACTTCCGCAAGTACACGCGCTTCAACATCCCAAGATCAAC 1080
Db 1021 AGCGTCCCCCTGAGCAAGAGACTTCCGCAAGTACACGCGCTTCAACATCCCAAGATCAAC 1080
Qy 1081 AACGAGACCCCGGCGCATCCGCTACAGTACAGTGTGCTGCCAAGGCTGAGAGGAGGAC 1140
Db 1081 AACGAGACCCCGGCGCATCCGCTACAGTGTGCTGCCAAGGCTGAGAGGAGGAC 1140
Qy 1141 CCGAGCATCTTCCAGAGACGATGACCAAGATCTTGAGAGCCCTTCCGCGCGCCCAACCCC 1200
Db 1141 CCGAGCATCTTCCAGAGACGATGACCAAGATCTTGAGAGCCCTTCCGCGCGCCCAACCCC 1200
Qy 1201 GAGATGTGATCTACCAAGGCCCCCTGTGATGAGGACGACCTGAGATCGGACGACAC 1260
Db 1201 GAGATGTGATCTACCAAGGCCCCCTGTGATGAGGACGACCTGAGATCGGACGACAC 1260
Qy 1261 CCGGCAAGATCGAGAGAGTGGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGAC 1320
Db 1261 CCGGCAAGATCGAGAGAGTGGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGAC 1320
Qy 1321 AAGAGACCAAGAGAGAGCCCCCTTCTGTGATGAGGCTACAGAGCTGCAACCCCGACAAG 1380
Db 1321 AAGAGACCAAGAGAGAGCCCCCTTCTGTGATGAGGCTACAGAGCTGCAACCCCGACAAG 1380
Qy 1381 TGGACCGTGAAGCCCATCGAGCTGCCCCGAGAGAGAGCTGACCCGTGAACGACATTCAG 1440
Db 1381 TGGACCGTGAAGCCCATCGAGCTGCCCCGAGAGAGAGCTGACCCGTGAACGACATTCAG 1440
Qy 1441 AAGCTGTGGGCAAGCTGAACCTGGGCGAGCAAGATCTAACCCCGGCAATCAAGGTGGCGAG 1500
Db 1441 AAGCTGTGGGCAAGCTGAACCTGGGCGAGCAAGATCTAACCCCGGCAATCAAGGTGGCGAG 1500
Qy 1501 CTGTGAAGCTGTGCGCGGCGCAAGGCGCTGACCGACATCGTGCCTTGAACCGAGAG 1560
Db 1501 CTGTGAAGCTGTGCGCGGCGCAAGGCGCTGACCGACATCGTGCCTTGAACCGAGAG 1560
Qy 1561 GCCGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGCGAGCCCGTGTCAACGCGGTGAC 1620
Db 1561 GCCGAGCTGAGCTGCGCGAGAACCGGAGATCTGTGCGAGCCCGTGTCAACGCGGTGAC 1620
Qy 1621 TAGGACCCCGAGAGAGACTGTGTGGCCGAGATCCAGAGAGAGGCGCACGACGTGAGCC 1680
Db 1621 TAGGACCCCGAGAGAGACTGTGTGGCCGAGATCCAGAGAGAGGCGCACGACGTGAGCC 1680

Db 1621 TAGGACCCCGAGAGAGACTGTGTGGCCGAGATCCAGAGAGAGGCGCACGACGTGAGCC 1680
Qy 1681 TACGAGATTTACCAAGAGAGGCTTCAAGAACCTGGAAGACCGGCAAGTACCCCAAGTGGCC 1740
Db 1681 TACGAGATTTACCAAGAGAGGCTTCAAGAACCTGGAAGACCGGCAAGTACCCCAAGTGGCC 1740
Qy 1741 ACCGCGCAACCAAGAGAGAGAGCTGACCGAGGCGGTGACAGAGATTCGCAATGAG 1800
Db 1741 ACCGCGCAACCAAGAGAGAGAGCTGACCGAGGCGGTGACAGAGATTCGCAATGAG 1800
Qy 1801 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCGCTGCGCTGACCAAGAGAGACTTGG 1860
Db 1801 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCGCTGCGCTGACCAAGAGAGACTTGG 1860
Qy 1861 GAGACTGTGAGACCGATCTAGGAGGCGACCTGATCCCGAGTGGAGGTTCTGTGAAC 1920
Db 1861 GAGACTGTGAGACCGATCTAGGAGGCGACCTGATCCCGAGTGGAGGTTCTGTGAAC 1920
Qy 1921 ACCCCCCCTGTGAGAGCTGTGTGTACAGCTGAGAGAGAGGCCCATCAATCGGCGCGAG 1980
Db 1921 ACCCCCCCTGTGAGAGCTGTGTGTGTACAGCTGAGAGAGAGGCCCATCAATCGGCGCGAG 1980
Qy 1981 ACCCTTCAAGTGAACCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACTG 2040
Db 1981 ACCCTTCAAGTGAACCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTACTG 2040
Qy 2041 ACCGACCGGCGCGCGCAAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGCGCGCGCAAGAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGACCGAG 2100
Qy 2101 CTGCGAGCCATCAAGTGGCCCTGCGAGGCAAGGCGAGGAGGTGAACATCGTGAACGAG 2160
Db 2101 CTGCGAGCCATCAAGTGGCCCTGCGAGGCAAGGCGAGGAGGTGAACATCGTGAACGAG 2160
Qy 2161 AGCAGATAGCCCTGCGCATCATTCAGGCGCCAGCCGCAAGAGAGAGAGAGAGCTGTG 2220
Db 2161 AGCAGATAGCCCTGCGCATCATTCAGGCGCCAGCCGCAAGAGAGAGAGAGAGCTGTG 2220
Qy 2221 AACGAGTATGAGAGAGAGCTTCAAGAGAGAGAGAGAGTGTACTTGAAGCTGGGCGGCC 2280
Db 2221 AACGAGTATGAGAGAGAGAGCTTCAAGAGAGAGAGAGAGTGTACTTGAAGCTGGGCGGCC 2280
Qy 2281 CACAGAGGATGCGCGCGCAACGAGAGATGACAACTGTGTGAGAGAGGCGATCCGCAAG 2340
Db 2281 CACAGAGGATGCGCGCGCAACGAGAGATGACAACTGTGTGAGAGAGGCGATCCGCAAG 2340
Qy 2341 GTGCTGTCTGTGAGCGGCAATCGATGCGGCGATGTGATCTACAGATACATGAGAGACTG 2400
Db 2341 GTGCTGTCTGTGAGCGGCAATCGATGCGGCGATGTGATCTACAGATACATGAGAGACTG 2400
Qy 2401 TACGTGGGCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCTAGACCGGT 2457
Db 2401 TACGTGGGCAAGCGGCGGCGCTAGAGATGATTAAGCTTCCGCGGCTAGACCGGT 2457
RESULT 5
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; env; nef;
KW immunostimulant; gene therapy; gene; de.
OS Human immunodeficiency virus; type C.
XX
PN Synthetic.
XX
PN W0200204493-A2.

Db 361 GGACACCTTGAACTTCCCCCAGATCACTCTGTGAGACGCCCCCTGTGTGATCAAGGTG 420
Qy 421 GGCGGCGAGATCAAGAGAGGCGCTGTGACACCGGCGCGAGACACCGTGTGAGAG 480
Db 421 GGCGGCGAGATCAAGAGAGGCGCTGTGACACCGGCGCGAGACACCGTGTGAGAG 480
Qy 481 ATGAGCCTTGCGCGAGAGTGAAGCCCAAGATGATCGCGGAGTCCGCGCTTCAACAG 540
Db 481 ATGAGCCTTGCGCGAGAGTGAAGCCCAAGATGATCGCGGAGTCCGCGCTTCAACAG 540
Qy 541 GTGCGGCGAGATCAAGACCAATCTGTATCGAGATCTGCGGAGAGAGGCGCATCGGCGAG 600
Db 541 GTGCGGCGAGATCAAGACCAATCTGTATCGAGATCTGCGGAGAGAGGCGCATCGGCGAG 600
Qy 601 CTGATCGGCGCGCGCGCGTGAACATCATCGGCGCGACATCTGACCGCTGAGCTGCG 660
Db 601 CTGATCGGCGCGCGCGCGTGAACATCATCGGCGCGACATCTGACCGCTGAGCTGCG 660
Qy 661 ACCCTGAATCTTCCCATCAGCGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
Db 661 ACCCTGAATCTTCCCATCAGCGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGAGT 720
Qy 721 GACGCGCGCGAGAGTGAAGAGTGGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
Db 721 GACGCGCGCGAGAGTGAAGAGTGGCGCTGTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
Qy 781 ATCTGAGAGAGATGAGAGAGAGGCGAGATCAACAGATCGGCGCGGAGAGAGCCCTTAC 840
Db 781 ATCTGAGAGAGATGAGAGAGAGGCGAGATCAACAGATCGGCGCGGAGAGAGCCCTTAC 840
Qy 841 AACACCCCGGTGTGCGCATCAAGAGAGAGACAGACCAAGTGGCGAGCTGTGTGAC 900
Db 841 AACACCCCGGTGTGCGCATCAAGAGAGAGACAGACCAAGTGGCGAGCTGTGTGAC 900
Qy 901 TTCCGCGAGCTGAACAAGCGACCCAGAGACTTCTGTGAGGTGCAAGTGGCGATCCCGGAC 960
Db 901 TTCCGCGAGCTGAACAAGCGACCCAGAGACTTCTGTGAGGTGCAAGTGGCGATCCCGGAC 960
Qy 961 CCCGCGCGCTGAAGAGAGAGAGCGGTGACCGTGTGAGAGTGGCGAGCGCTTACTTTC 1020
Db 961 CCCGCGCGCTGAAGAGAGAGAGCGGTGACCGTGTGAGAGTGGCGAGCGCTTACTTTC 1020
Qy 1021 AGCGTGCCTGTGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGATCAAC 1080
Db 1021 AGCGTGCCTGTGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCGAGATCAAC 1080
Qy 1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAGTGTGTGCGCCAGGAGCTGAGAGGCGAGC 1140
Db 1081 AACGAGACCCCGGAGATCCGCTTACCAAGTACAGTGTGTGCGCCAGGAGCTGAGAGGCGAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGAGAGTGAACCAAGATCTGTGAGCGCTTCCGCGCGCGAGACCC 1200
Db 1141 CCCAGCATCTTCCAGAGAGAGTGAACCAAGATCTGTGAGCGCTTCCGCGCGCGAGACCC 1200
Qy 1201 GAGATCTGTATTTACCA-----GGCCCCCTGTATGTGTGAGAGCGACTGTGAGATCGG 1254
Db 1201 GAGATCTGTATTTACCA-----GGCCCCCTGTATGTGTGAGAGCGACTGTGAGATCGG 1254
Qy 1255 CAGGACCGGCGCAAGTGAAGAGAGTGTGCGAGACCTGTGCGCGCTTCAACAGC 1314
Db 1255 CAGGACCGGCGCAAGTGAAGAGAGTGTGCGAGACCTGTGCGCGCTTCAACAGC 1314
Qy 1261 CAGGACCGGCGCAAGTGAAGAGAGTGTGCGAGACCTGTGCGCGCTTCAACAGC 1320
Db 1261 CAGGACCGGCGCAAGTGAAGAGAGTGTGCGAGACCTGTGCGCGCTTCAACAGC 1320
Qy 1315 CCCGACAGAGAGACCAAGAGAGCGCGCTTCTGTGAGTGGGCTGAGAGCTGAGACCC 1374
Db 1315 CCCGACAGAGAGACCAAGAGAGCGCGCTTCTGTGAGTGGGCTGAGAGCTGAGACCC 1374
Qy 1375 GACAGAGTGAACCGTGTGAGCGCGCATCGAGCTGCGCGAGAGAGAGAGTGTGAGACG 1434
Db 1375 GACAGAGTGAACCGTGTGAGCGCGCATCGAGCTGCGCGAGAGAGAGAGTGTGAGACG 1434
Qy 1381 GACAGAGTGAACCGTGTGAGCGCGCATCGAGCTGCGCGAGAGAGAGAGTGTGAGACG 1440
Db 1381 GACAGAGTGAACCGTGTGAGCGCGCATCGAGCTGCGCGAGAGAGAGAGTGTGAGACG 1440
Qy 1435 ATTCGAGAGCTGTGTGAGAGAGTGAAGTGTGAGCGAGCGAGATTTACCGCGGATCAAGGTG 1494
Db 1435 ATTCGAGAGCTGTGTGAGAGAGTGAAGTGTGAGCGAGCGAGATTTACCGCGGATCAAGGTG 1494
Qy 1441 ATTCGAGAGCTGTGTGAGAGAGTGAAGTGTGAGCGAGCGAGATTTACCGCGGATCAAGGTG 1500
Db 1441 ATTCGAGAGCTGTGTGAGAGAGTGAAGTGTGAGCGAGCGAGATTTACCGCGGATCAAGGTG 1500

Qy 1495 CGCCAGCTGTGAAGCTGTGCGCGCGCGCAAGGCGCTGACCGAGCATGTGCGCTTGAAC 1554
Db 1501 CGCCAGCTGTGAAGCTGTGCGCGCGCGCGCAAGGCGCTTGAACCGAGCATGTGTGCGCTTGAAC 1560
Qy 1555 GAGAGGCGAGAGCTGTGAGCTGTGCGCGAGAACCGCGAGATCTGTGCGGAGCGCTGTGACGCG 1614
Db 1561 GAGAGGCGAGAGCTGTGAGCTGTGCGCGAGAACCGCGAGATCTGTGCGGAGCGCGCTGTGACGCG 1620
Qy 1615 GTGTACTAGACCCCGAGAGAGAGCTGTGTGCGCGAGATCTGAGAGAGAGAGAGAGAGAGAG 1674
Db 1621 GTGTACTAGACCCCGAGAGAGAGCTGTGTGCGCGAGATCTGAGAGAGAGAGAGAGAGAGAG 1680
Qy 1675 TGAGCTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGAGAGTACGCGCAAG 1734
Db 1681 TGAGCTTACAGATCTTACAGAGAGCGCTTCAAGAGAGCTGAGAGAGCGGAGAGTACGCGCAAG 1740
Qy 1735 ATGCGGACCGCGCGACCAAG 1794
Db 1741 ATGCGGACCGCGCGACCAAG 1800
Qy 1795 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1854
Db 1801 ATGAGAGAGATCTGTATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1855 ACTTGGAGAGAGCTGTGTGAG 1914
Db 1861 ACTTGGAGAGAGCTGTGTGAG 1920
Qy 1915 GTGAACACCCCGCTGTGTGAAGTGTGTGAACAGTGTGAGAGAGAGAGAGAGAGAGAGAG 1974
Db 1921 GTGAACACCCCGCTGTGTGAAGTGTGTGAACAGTGTGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1975 GCGGAGAGCTTCTTATGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2034
Db 1981 GCGGAGAGCTTCTTATGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy 2035 TACGTGAG 2094
Db 2041 TACGTGAG 2100
Qy 2095 ACCGAGCTGAG 2154
Db 2101 ACCGAGCTGAG 2160
Qy 2155 ACCGAG 2214
Db 2161 ACCGAG 2220
Qy 2215 CTGTGTGAACAGATCATGAG 2274
Db 2221 CTGTGTGAACAGATCATGAG 2280
Qy 2275 CCCGCGCGAG 2334
Db 2281 CCCGCGCGAG 2340
Qy 2335 CGCAG 2394
Db 2341 CGCAG 2400
Qy 2395 GACCTGTAG 2454
Db 2401 GACCTGTAG 2460
Qy 2455 GGTGAATTC 2463
Db 2461 GGTGAATTC 2469

RESULT 7
ABI3961
ID ABI3961 standard; DNA; 2457 BP.

XX ABL39961;
AC
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975YMMN SEQ ID NO:32.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; da.
XX
OS Human immunodeficiency virus; type C.
OS Synthetic.
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
XX (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;
XX
DR WPI; 2002-154920/20.
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunisation or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 10; 233pp; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (1). (i) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other:

Query Match 98.9%; Score 2436.2; DB 6; Length 2457;
Best Local Similarity 99.6%; Pred. No. 9.3e-294;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GTTGAAGCCACCTTGGCCGAGGCGCATGAGCCAGGCCACCGCCCAACATCTGATGACG 60
DB 1 GTGAGCGCCACCTATGGCGGAGCCATGAGCCAGGCCACCGCCCAACATCTGATGACG 60
QY 61 CGAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120
DB 61 CGAGCAACTTCAAGGGGCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGGC 120
QY 121 CACATGCGCCCGCACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGAG 180
DB 121 CACATGCGCCCGCACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGAG 180
QY 121 CACATGCGCCCGCACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGAG 180
DB 121 CACATGCGCCCGCACTGCGCGCGCCCGCCGCAAGAGGGCTGTGGAAGTGCGGCAAGAG 180
QY 181 GGGCAACCAATGAAGAGCTGCAACGAGCGCCAACTTCTTCCGCGAGACCTGGCC 240
DB 181 GGGCAACCAATGAAGAGCTGCAACGAGCGCCAACTTCTTCCGCGAGACCTGGCC 240
QY 241 TTCCCGGAGGCAAGGCGCGGAGTTCCCGAGCAGAGAAACCGCGCAACGCGCCAC 300
DB 241 TTCCCGGAGGCAAGGCGCGGAGTTCCCGAGCAGAGAAACCGCGCAACGCGCCAC 300

QY 301 AGCCGAGCTGACAGTGTGCGCGCAGCAACCCCGCAGCGAGCGCGCCGAGCGCCAG 360
DB 301 AGCCGAGCTGACAGTGTGCGCGCAGCAACCCCGCAGCGAGCGCGCCGAGCGCCAG 360
QY 361 GGCACCTTGAATCTTCCCGCAATCACTGTGTGAGCGCGCCCTGTGTGAGCATCAAGTGTG 420
DB 361 GGCACCTTGAATCTTCCCGCAATCACTGTGTGAGCGCGCCCTGTGTGAGCATCAAGTGTG 420
QY 421 GCGCGGCAATGATGAGAGAGCGCTGTGTGACACCGCGCGCAAGCAACCGTGTGAGAGAG 480
DB 421 GCGCGGCAATGATGAGAGAGCGCTGTGTGACACCGCGCGCAAGCAACCGTGTGAGAGAG 480
QY 481 ATGAGCTGCGCGGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
DB 481 ATGAGCTGCGCGGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGCTTCAATCAAG 540
QY 541 GTGCGCCAGTACGACAGATCTGTATCGATCTGCGGCAAGAGCCATCGGCAACCGTG 600
DB 541 GTGCGCCAGTACGACAGATCTGTATCGATCTGCGGCAAGAGCCATCGGCAACCGTG 600
QY 601 CTGATCGGCGCCCAACCCCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGCTTCG 660
DB 601 CTGATCGGCGCCCAACCCCGTGAACATCATCGCGCGCAACATCTGACCCAGCTGGCTTCG 660
QY 661 ACCCTGAATCTTCCCATGACGCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
DB 661 ACCCTGAATCTTCCCATGACGCCCATCGAGACCGTGCCCGTGAAGCTGAAGCCCGGCATG 720
QY 721 GACGCGCCCAAGTGAAGAGTGTGCGCCCTGACCGGAGAGAAATCAAGGCCCTGACCGGC 780
DB 721 GACGCGCCCAAGTGAAGAGTGTGCGCCCTGACCGGAGAGAAATCAAGGCCCTGACCGGC 780
QY 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACAGATGCGCCCGAGAACCTCTAC 840
DB 781 ATCTGAGAGAGATGAGAGAGGCGCAAGATCAACAGATGCGCCCGAGAACCTCTAC 840
QY 841 AACACCCCGTGTGTCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
DB 841 AACACCCCGTGTGTCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 901 TTCGCGAGCTGAACAGAGGCAACCGAGACTTCTGTGAGGTGACGTGCGCATCCCGCAC 960
DB 901 TTCGCGAGCTGAACAGAGGCAACCGAGACTTCTGTGAGGTGACGTGCGCATCCCGCAC 960
QY 961 CCGCGCGGCTGAAG 1020
DB 961 CCGCGCGGCTGAAG 1020
QY 1021 AGCGTGCCTTGAAG 1080
DB 1021 AGCGTGCCTTGAAG 1080
QY 1081 AACGAGAGCCCGGCGCATCGCTACAGTACCAAGTCTGTGAGCGCGGCTGAGAGGGCAGC 1140
DB 1081 AACGAGAGCCCGGCGCATCGCTACAGTACCAAGTCTGTGAGCGCGGCTGAGAGGGCAGC 1140
QY 1141 CCCAGCATCTTCAAG 1200
DB 1141 CCCAGCATCTTCAAG 1200
QY 1201 GAGATGTGATCTTCAAG 1260
DB 1201 GAGATGTGATCTTCAAG 1260
QY 1261 CGGCGCAAGTGAAG 1320
DB 1261 CGGCGCAAGTGAAG 1320
QY 1321 AAGAGACACCAAG 1380
DB 1321 AAGAGACACCAAG 1380
QY 1374 AAGAGACACCAAG 1374
DB 1374 AAGAGACACCAAG 1374
QY 1440 TGAACCGTGAAG 1440
DB 1440 TGAACCGTGAAG 1440

```

Db      1375  TGGACCGTGCAGCCCATCGAGCTGCCCGGAGGAGGAGCTGGAACGTGAACGATCCAG
Qy      1441  AAGCTGTGTGGGAAAGCTGAAGTGGGCGACCGAGATCTAACCCGGGATCGAAGGTGGCCGAG
Db      1435  AAGCTGTGTGGGAAAGCTGAAGTGGGCGACCGAGATCTAACCCGGGATCGAAGGTGGCCGAG
Qy      1501  CTGTGCAAGCTGTGTGGGCGCCGCAAGGCTTGAACGAGCATGTGCTCCCTGACCGAGGAG
Db      1495  CTGTGCAAGCTGTGTGGGCGCCGCAAGGCTTGAACGAGCATGTGCTCCCTGACCGAGGAG
Qy      1561  GCCGAGCTGTGAGCTGGCCGAGAACCGCGAGATCTTCCGCGAGCCGTGCAAGCGCTGTAC
Db      1555  GCCGAGCTGTGAGCTGGCCGAGAACCGCGAGATCTTCCGCGAGCCGTGCAAGCGCTGTAC
Qy      1621  TACGAGCCCGGAGAGGAGCTGTGTGGCGGAGATCCGAGAACGAGGCGACGACGATGGAGCC
Db      1615  TACGAGCCCGGAGAGGAGCTGTGTGGCGGAGATCCGAGAACGAGGCGACGACGATGGAGCC
Qy      1681  TACGAGATCTTACGAGAGGCTTCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGC
Db      1675  TACGAGATCTTACGAGAGGCTTCTTCAAGAACTTGAAGACCGGCAAGTACGCAAGATGCGC
Qy      1741  ACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAAGATCCGCAATGAG
Db      1735  ACCGCCCAACCAACGACGTGAAGCAGCTGACCGAGGCGGTGCAAGATCCGCAATGAG
Qy      1801  AACATCTGTATCTGTGGGCGAAGACCCCGAAGTTCCGCTGCGCATCCAGAAAGAGACTGG
Db      1795  AACATCTGTATCTGTGGGCGAAGACCCCGAAGTTCCGCTGCGCATCCAGAAAGAGACTGG
Qy      1861  GAGACCTGTGTGAGCCGACTACTGTGAGCGCCACCTGTGATCCCGAGTGGAGTTCGTGAAC
Db      1855  GAGACCTGTGTGAGCCGACTACTGTGAGCGCCACCTGTGATCCCGAGTGGAGTTCGTGAAC
Qy      1921  ACCCCCCCTGTGTGAAAGCTGTGTGTAACAGCTGTGAGAGAGAGCCATCATTCGCGCCGAG
Db      1915  ACCCCCCCTGTGTGAAAGCTGTGTGTAACAGCTGTGAGAGAGAGCCATCATTCGCGCCGAG
Qy      1981  ACCTTCTAGTGTGAGCGGCGCCCGCAACCGCGGAGACCAAGATGGGAGCGCGCTACGCG
Db      1975  ACCTTCTAGTGTGAGCGGCGCCCGCAACCGCGGAGACCAAGATGGGAGCGCGCTACGCG
Qy      2041  ACCGACCGGCGCGCGGCAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGCCGAG
Db      2035  ACCGACCGGCGCGCGGCAAGATCTGTGAGCTTGAACCGAGACCAACCAAGAGCCGAG
Qy      2101  CTGCAAGGCGCATCGAGCTGGCCCTTGCAGAGCAACCGCGAGCGAGGTGAACATCTGTGACGAC
Db      2095  CTGCAAGGCGCATCGAGCTGGCCCTTGCAGAGCAACCGCGAGCGAGGTGAACATCTGTGACGAC
Qy      2161  AACCGATGAGCGCTGTGGGCGATCTCCAGGCGCCGAGCCGCAAGAGCGAGAGCGAGCTGGT
Db      2155  AACCGATGAGCGCTGTGGGCGATCTCCAGGCGCCGAGCCGCAAGAGCGAGAGCGAGCTGGT
Qy      2221  AACCGATGATCGAGGAGCTGATCAAGAGAGAGAGGTGTACTGTGAGCTGGGTGCCGCC
Db      2215  AACCGATGATCGAGGAGCTGATCAAGAGAGAGAGGTGTACTGTGAGCTGGGTGCCGCC
Qy      2281  CACAAAGGCGATCGGCGGCAACGAGCATGACCAAGCTGTGTGAGCAAGGCGATCCGCAAG
Db      2275  CACAAAGGCGATCGGCGGCAACGAGCATGACCAAGCTGTGTGAGCAAGGCGATCCGCAAG
Qy      2341  GTGCTGTCTTCTGAGACGGGCGATCGATGGCGGCGATCTGTGATCTTACAGATCACTGAGCACTG
Db      2335  GTGCTGTCTTCTGAGACGGGCGATCGATGGCGGCGATCTGTGATCTTACAGATCACTGAGCACTG
Qy      2401  TACGTGTGAGGCGCGCGCTAGAGATCGATTAAAGCTTCCCGGCGCTAGCAACCGGTGAA
Db      2395  TACGTGTGAGGCGCGCGCTAGAGATCGATTAAAGCTTCCCGGCGCTAGCAACCGGTGAA
Qy      2461  TTC 2463

```

```

Db      2455  TTC 2457
Qy      121  GTCGACGCGCAACATGCGCGAGGCGCATGAGCGGCGCAACGCGCAACATCTGTGATGCGAG
Db      120  GTCGACGCGCAACATGCGCGAGGCGCATGAGCGGCGCAACGCGCAACATCTGTGATGCGAG
Qy      61  CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGAGGCGC
Db      60  CGCAGCAACTTCAAGGCGCCCAAGCGCATCATCAAGTGTCTTCAACTGCGCGCAAGAGGCGC
Qy      121  CACATGCGCGCGCAACTGCGCGCGCCCGCAAGAGAGGCTGTGTGAGAGTGTGCGCAAGAG
Db      120  CACATGCGCGCGCAACTGCGCGCGCCCGCAAGAGAGGCTGTGTGAGAGTGTGCGCAAGAG
Qy      180  121  CACATGCGCGCGCAACTGCGCGCGCCCGCAAGAGAGGCTGTGTGAGAGTGTGCGCAAGAG
Db      180  121  CACATGCGCGCGCAACTGCGCGCGCCCGCAAGAGAGGCTGTGTGAGAGTGTGCGCAAGAG

RESULT 8
ADM73766 standard; DNA; 2457 BP.
ID ADM73766;
AC ADM73766;
XX 03-JUN-2004 (first entry)
XX
XX HIV-1 polynucleotide #9.
DE HIV-1; gene; de; HIV pol; immune response; DNA immunisation;
XX HIV type C protein; immunostimulant.
XX
XX Human immunodeficiency virus 1.
OS
XX US200323961-A1.
XX
XX 04-DEC-2003.
XX
XX 05-JUL-2001; 2001US-00899575.
XX
XX 05-JUL-2000; 2000US-00610313.
XX
XX (MGRG/) MGRGDR J Z.
XX (BARN/) BARNETT S W.
XX (ENG/) ENGELBRECHT S.
XX (RENS/) RENSBERG B J V.
XX
XX Megede JZ, Barnett SW, Engelbrecht S, Rensburg BJV;
XX WPI; 2004-060515/06.
XX
XX New expression cassette comprising a polynucleotide sequence encoding an
XX HIV Pol polypeptide, useful in eliciting an immune response, in DNA
XX immunisation, generating of packaging cell lines or in producing HIV Type
XX C proteins.
XX
XX Claim 1; SEQ ID NO 32; 160pp; English.
XX
XX The invention relates to an expression cassette comprising a
XX polynucleotide sequence encoding an HIV Pol polypeptide. The invention
XX also relates to a recombinant expression system for use in a host cell
XX comprising an expression cassette, where the polynucleotide sequence
XX further comprises control elements capable of driving expression in the
XX selected host cell, a cell comprising an expression cassette where the
XX polynucleotide sequence further comprises control elements compatible
XX with the expression in the cell and a composition for generating an
XX immunological response, comprising an expression cassette. The expression
XX cassette and the methods of the invention are useful in eliciting an
XX immune response, in DNA immunisation, in generation of packaging cell
XX lines and in producing HIV Type C proteins. This sequence represents an
XX HIV-1 polynucleotide of the invention.
XX
XX Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;
XX
Query Match 98.9%; Score 2436.2; DB 12; Length 2457;
Best Local Similarity 99.6%; Pred. No. 9.3e-294;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

```


OY	181	GGCCACCAAGATGAAGACTGCAACGAGCGCAAGCCCAATTCTTCCGCGAGAGACTGAGCC	240
Db	181	GGCCACCAAGATGAAGACTGCAACGAGCGCGCAAGCCCAATTCTTCCGCGAGAGACTGAGCC	240
OY	241	TTCCCCCAAGGSCAAGSCCCCGGAGATTCCCGACGAGAGAACCGGCGCAACAGCCCAACC	300
Db	241	TTCCCCCAAGGSCAAGSCCCCGGAGATTCCCGACGAGAGAACCGGCGCAACAGCCCAACC	300
OY	301	AGCCGCGAGCTGCAGAGTGCAGGCGGCAACCCCGCAGCGAGGCGCGCGCGCGCGCAG	360
Db	301	AGCCGCGAGCTGCAGAGTGCAGGCGGCAACCCCGCAGCGAGGCGCGCGCGCGCGCAG	360
OY	361	GGCAACCTTGAACTTCCCGCCAAATCACTCCCTGGGAGAGGCCCCCGTGGTGAATCAAAAGTG	420
Db	361	GGCAACCTTGAACTTCCCGCCAAATCACTCCCTGGGAGAGGCCCCCGTGGTGAATCAAAAGTG	420
OY	421	GGCGGCGAGATCAAGAGAGGCGCTGCTGGACAACCGGCGCGACGACACCGTGTGGAGAG	480
Db	421	GGCGGCGAGATCAAGAGAGGCGCTGCTGGACAACCGGCGCGACGACACCGTGTGGAGAG	480
OY	481	ATGAGCCTGCGCCGCAAGTGAAGCCCAAGATGATCGGCGGATCGCGGCTTCATCAAG	540
Db	481	ATGAGCCTGCGCCGCAAGTGAAGCCCAAGATGATCGGCGGATCGCGGCTTCATCAAG	540
OY	541	GTGCGCGCAGTACAGACCAATCTGTATGAGATCTGCGCGCAAGAGGCAATCGGACCGTG	600
Db	541	GTGCGCGCAGTACAGACCAATCTGTATGAGATCTGCGCGCAAGAGGCAATCGGACCGTG	600
OY	601	CTGATCGAGCCCCCAACCCCGGTGAACATCATCGGCGCGCAACATGCTGAGCCGACTGGCTGC	660
Db	601	CTGATCGAGCCCCCAACCCCGGTGAACATCATCGGCGCGCAACATGCTGAGCCGACTGGCTGC	660
OY	661	ACCTTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCAATG	720
Db	661	ACCTTGAACTTCCCATCAGCCCATCGAGACCGTGCCTGGAAGCTGAAGCCCGGCAATG	720
OY	721	GACGCGCCCCCAAGGTGAACGATGCGCCCCCTGACCCAGAGAAAGTCAAGGCGCTTACCGCGC	780
Db	721	GACGCGCCCCCAAGGTGAACGATGCGCCCCCTGACCCAGAGAAAGTCAAGGCGCTTACCGCGC	780
OY	781	ATCTCGAGAGATGAGAAAGAGGGGCAAGATCAACAAATCGGCGCCCGAGAACCCCTTAC	840
Db	781	ATCTCGAGAGATGAGAAAGAGGGGCAAGATCAACAAATCGGCGCCCGAGAACCCCTTAC	840
OY	841	AAACAACCCCGTGTTCGCATCAAGAAAGAGCAGACCAAGTGGCGCAAGCTGTGTGAC	900
Db	841	AAACAACCCCGTGTTCGCATCAAGAAAGAGCAGACCAAGTGGCGCAAGCTGTGTGAC	900
OY	901	TTCCGCGAGCTGAACAAGCGCACCGAGAACTTCTGGAGAGTGCAGCTGGGCAATCCCGCAC	960
Db	901	TTCCGCGAGCTGAACAAGCGCACCGAGAACTTCTGGAGAGTGCAGCTGGGCAATCCCGCAC	960
OY	961	CCGCGCGGCGCTGAAGAGAAAGAGAGCGTGAACGTGTGGAAGTGGGCGAGCGCTTACTTC	1020
Db	961	CCGCGCGGCGCTGAAGAGAAAGAGAGCGTGAACGTGTGGAAGTGGGCGAGCGCTTACTTC	1020
OY	1021	AGCGTGCCTTCGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCAAGATCAAC	1080
Db	1021	AGCGTGCCTTCGAGAGAGACTTCCGCAAGTACACCGCTTTCACCATCCCAAGATCAAC	1080
OY	1081	AACGAGACCCCGCGGCAATCGCTACCAAGTCAAAAGTGTGCTGCCAAGGCTGGAAGGGCAGC	1140
Db	1081	AACGAGACCCCGCGGCAATCGCTACCAAGTCAAAAGTGTGCTGCCAAGGCTGGAAGGGCAGC	1140
OY	1141	CCCAAGATTTTCAAGAGCAGATGACAAAGATCTGGAGCCCTTTCGCGCGCGCAACCC	1200
Db	1141	CCCAAGATTTTCAAGAGCAGATGACAAAGATCTGGAGCCCTTTCGCGCGCGCAACCC	1200
OY	1201	GAGATGTGATCTTACCAAGGCGCCCGCTGTACGTGGGCAAGCACTTGAGATCGGCGCAGCAGC	1260
Db	1201	GAGATGTGATCTTACCAAGGCGCCCGCTGTACGTGGGCAAGCACTTGAGATCGGCGCAGCAGC	1260

QY	1261	CGCCCAAGATCGAGGAGCTGCCAAGCACTCTCTGGCGTGGGGCTTCAACACCCCGAC	1320
Db	1261	CGCCCAAGATCGAGGAGCTGCCAAGCACTCTCTGGCGTGGGGCTTCAACACCCCGAC	1320
QY	1321	AAGAAAGCAACGAAGAGGCCCTTCTCTGTGATGGGGCTTACAGAGCTGCACCCCGACAAG	1380
Db	1321	AAGAAAGCAACGAAGAGGCCCTTCTCTGTGATGGGGCTTACAGAGCTGCACCCCGACAAG	1380
QY	1381	TGAGCCGTGCAGCCCATCGAGCTGCCCGAGAAGAGAGCTGAGCCGTGAACGATCTCAG	1440
Db	1375	TGAGCCGTGCAGCCCATCGAGCTGCCCGAGAAGAGAGCTGAGCCGTGAACGATCTCAG	1434
QY	1441	AAGCTGTGGGCAAGCTTGAACTGGGCGACCCAGATCTTACCCCGCATCAGAGTGGCGCAG	1500
Db	1435	AAGCTGTGTGGGCAAGCTTGAACTGGGCGACCCAGATCTTACCCCGCATCAGAGTGGCGCAG	1494
QY	1501	CTGTGCAGAGCTGTGGCGGCGCGCAAGGCCCTGACCGACATCTGTGCCCTTGACCGAAGAG	1560
Db	1495	CTGTGCAGAGCTGTGGCGGCGCGCAAGGCCCTTGAACCGACATCTGTGCCCTTGACCGAAGAG	1554
QY	1561	GCCGAGCTGAGAGCTGGCCGAGAACCGCGAGATCTTGCGCGAGCCCGTGCACGGCGTGTAC	1620
Db	1555	GCCGAGCTGAGAGCTGGCCGAGAACCGCGAGATCTTGCGCGAGCCCGTGCACGGCGTGTAC	1614
QY	1621	TACGACCCCGACGAGAGACTGTGTGGCCGAGATTCAGAAAGCAAGGCGCACGACTGACCC	1680
Db	1615	TACGACCCCGACGAGAGACTGTGTGGCCGAGATTCAGAAAGCAAGGCGCACGACTGACCC	1674
QY	1661	TACCGAGATCTTACCGAGGAGCCCTTCAAGACCTGGAAGA CCGGCAAGTACCGGCAAGTGC	1740
Db	1675	TACCGAGATCTTACCGAGGAGCCCTTCAAGAACTTGAAAGCCGGCAAGTACCGGCAAGTGC	1734
QY	1741	ACCGGCCACACCAAGAGAGTGAAGCACTGACCGAGGCCGTGCAGAAAGATCGCCATGAG	1800
Db	1735	ACCGGCCACACCAAGAGAGTGAAGCACTGACCGAGGCCGTGCAGAAAGATCGCCATGAG	1794
QY	1801	AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTCGTCCCATCCAGAAAGAGACTGG	1860
Db	1795	AGCATCTGTGATCTGGGGCAAGACCCCAAGTTCCGCTCGTCCCATCCAGAAAGAGACTGG	1854
QY	1861	GAGACTGTGTGAGATCCGACTTGTGGCAGAGCCACTGTGATCCCGAGTGGAGATTCTGTAA C	1920
Db	1855	GAGACTGTGTGAGATCCGACTTGTGGCAGAGCCACTGTGATCCCGAGTGGAGATTCTGTAA C	1914
QY	1921	ACCCCCCCCCCTGTGGAAGCTGTGGTACAAGCTGTGAGAAAGAGGCCCATCTTCGCGCCGAG	1980
Db	1915	ACCCCCCCCCCTGTGTGAAGCTGTGGTACAAGCTGTGAGAAAGAGGCCCATCTTCGCGCCGAG	1974
QY	1981	ACCTTCTTACTGTGAGCGGCGCGCCACCGCGAGACCAAGATCGGCAAGGCCGCGTACTGTC	2040
Db	1975	ACCTTCTTACTGTGAGCGGCGCGCCACCGCGAGACCAAGATCGGCAAGGCCGCGTACTGTC	2034
QY	2041	ACCGACCGGGGCGCGCAGAGAAATGTGTGAGCTTGACCGAGACCAACCAACGAGAAACCGAG	2100
Db	2035	ACCGACCGGGGCGCGCAGAGAAATGTGTGAGCTTGACCGAGACCAACCAACGAGAAACCGAG	2094
QY	2101	CTGCGAGGCATTCAGCTGAGCTGGCCCTGAGGACAGCGGCGACGAGGTGAACATCTGTACCGAC	2160
Db	2095	CTGCGAGGCATTCAGCTGAGCTGGCCCTGAGGACAGCGGCGACGAGGTGAACATCTGTACCGAC	2154
QY	2161	AGCCAGTACGCTCTGGGAGTATCCAGGCGCAGGCCCGACAAAGACGAGAGCGAGCTGGTC	2220
Db	2155	AGCCAGTACGCTCTGGGAGTATCCAGGCGCAGGCCCGACAAAGACGAGAGCGAGCTGGTC	2214
QY	2221	AACCAAGATCATGAGCAGCTGATCAAGAAAGAGAAAGTGTACTCTGAGCTGTGGTGCCTGCC	2280
Db	2215	AACCAAGATCATGAGCAGCTGATCAAGAAAGAGAAAGTGTACTCTGAGCTGTGGTGCCTGCC	2274
QY	2281	CACCAAGGGGATTCGGGCGAGACGAGGAGATTCGACAGAGCTGTGTGACCAAGGGGATCCGCAAG	2340
Db	2275	CACCAAGGGGATTCGGGCGAGACGAGGAGATTCGACAGAGCTGTGTGAGCAGAGGGGATCCGCAAG	2334
QY	2341	GTGCTGTCTTGGACGGGATCGATTCGATCTTACAGTACATGAGCGACTGTG	2400

Db	2335	GTCTGTTCTCTGGACGCGATCGATGGCGCATCGTGAATCACTACATGAGCAACTGG	2394
Qy	2401	TACGTGGCGACGGCGCGGCTTAGATCGATTAAAGCTTCCGGGGCTAGACCGGTGA	2460
Db	2395	TACGTGGCGACGGCGCGGCTTAGATCGATTAAAGCTTCCGGGGCTAGACCGGTGA	2455
Qy	2461	TTC 2463	
Db	2455	TTC 2457	
RESULT 9			
ID	ACA03548	standard; DNA; 2457 BP.	
XX	ACA03548;		
XX	22-MAY-2003	(first entry)	
XX	DT		
DE	Synthetic DNA encoding immunogenic HIV peptide #31.		
KW	Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;		
KW	gene therapy; packaging cell line; humoral immune response;		
KW	cellular immune response; gene delivery vector; DNA immunisation; de.		
XX	Synthetic.		
XX	WO2003004657-A1.		
XX	16-JAN-2003.		
XX	PD		
XX	05-JUL-2002; 2002WO-US021421.		
XX	05-JUL-2001; 2001US-0303192P.		
XX	31-AUG-2001; 2001US-031660P.		
XX	16-JAN-2002; 2002US-0349728P.		
XX	16-JAN-2002; 2002US-0349793P.		
XX	16-JAN-2002; 2002US-0349871P.		
XX	(CHIR) CHIRON CORP.		
XX	Zur Megede J, Barnett SW, Lian Y;		
XX	WPI; 2003-221602/21.		
XX	New synthetic polynucleotides encoding antigenic HIV type B and/or type C		
XX	polypeptides, useful as immunogenic compositions or vaccines for		
XX	generating humoral or cellular immune responses against HIV in a subject,		
XX	especially humans.		
XX	Example 1; Fig 36; 262pp; English.		
XX	The invention describes a synthetic polynucleotide encoding 2 or more		
XX	immunogenic HIV polypeptides, where at least 2 of the polypeptides are		
XX	derived from different HIV subtypes. The polynucleotide is useful for		
XX	immunisation, generation of packaging cell lines, or production of HIV		
XX	polypeptides. The polynucleotide and its encoded proteins are useful as		
XX	immunogenic compositions or vaccines for generating humoral or cellular		
XX	immune responses against HIV in a subject, or for inducing neutralising		
XX	antibodies against HIV. The gene delivery vector comprising the		
XX	polynucleotide is also useful for DNA immunisation of, or for generating		
XX	an immune response (e.g. a humoral or cellular immune response) in, a		
XX	subject such as a mammal, particularly a human. This sequence encodes a		
XX	human immunodeficiency virus immunogenic peptide		
XX	Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;		
XX	Query Match	98.7%; Score 2430.2; DB 8; Length 2457;	
XX	Best Local Similarity	99.6%; Pred. No. 5.2e-293;	
XX	Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1.		

Db	1	GCACACATGCGCAGGCGCATAGCCAGCCACACAGCGCAACATCTTGATGACAGCGAC	60
QY	67	AACCTCAAGGCGCCCAAGCGCATCATGAAGTGTCTTCAACTGCGGCAAGAGGCGCACATC	126
Db	61	AACCTCAAGGCGCCCAAGCGCATCATGAAGTGTCTTCAACTGCGGCAAGAGGCGCACATC	120
QY	127	GCCCCGCACTGCGCGCGCCCCCGCGCAAGAGGCGCTGTGAAATGCGCGCAAGAGGCGCAC	186
Db	121	GCCCCGCACTGCGCGCGCCCCCGCGCAAGAGGCGCTGTGAAATGCGCGCAAGAGGCGCAC	180
QY	187	CAGATGAAGGACTGCACTGAGGCGCAGGCGCAACTTCTTCGCGAGGACCTGAGCGCTTCCGC	246
Db	181	CAGATGAAGGACTGCACTGAGGCGCAGGCGCAACTTCTTCGCGAGGACCTGAGCGCTTCCGC	240
QY	247	CAGGCGCAAGCGCCCGAGTTCCCGACGAGCAAGAACCGCGCCAAACAGCCCAACAGCGCG	306
Db	241	CAGGCGCAAGCGCCCGAGTTCCCGACGAGCAAGAACCGCGCCAAACAGCCCAACAGCGCG	300
QY	307	GAGCTGACAGTGCAGCGCGCAACCCCGCGACGAGCGCGCGCGCGCGCGCACAGGCAACC	366
Db	301	GAGCTGACAGTGCAGCGCGCAACCCCGCGACGAGCGCGCGCGCGCGCGCACAGGCAACC	360
QY	367	CTGAACTTCCCCGAGATCACTCTGTGCGAGCGCCCCCTGTGAGACATCAAGAGTGGCGGC	426
Db	361	CTGAACTTCCCCGAGATCACTCTGTGCGAGCGCCCCCTGTGAGACATCAAGAGTGGCGGC	420
QY	427	CAGATCAAGAGGCGCTGTGCGACACGCGCGCGCACACCTGTGCTGAGAGATGAGC	486
Db	421	CAGATCAAGAGGCGCTGTGCGACACGCGCGCGCACACCTGTGCTGAGAGATGAGC	480
QY	487	CTGCCCCGCAAGTGAAGCGCCCAAGATGATCGCGCGCATTCGCGCTTTCATCAAGTGCAC	546
Db	481	CTGCCCCGCAAGTGAAGCGCCCAAGATGATCGCGCGCATTCGCGCTTTCATCAAGTGCAC	540
QY	547	CAGTACGACCAAGATCTGTATGGAATCTGCGCGCAAGAGGCGCATCGGCACCTGTCTGATC	606
Db	541	CAGTACGACCAAGATCTGTATGGAATCTGCGCGCAAGAGGCGCATCGGCACCTGTCTGATC	600
QY	607	GAGCCCAACCCCGGTGAACATCATGCGCGCGCAACATGCTGACCGAGCTGGGCTGCAACCTG	666
Db	601	GAGCCCAACCCCGGTGAACATCATGCGCGCGCAACATGCTGACCGAGCTGGGCTGCAACCTG	660
QY	667	AACCTCCCATCAGCCCCCATGAGACCGGTGCCTGTGAAGCTGAAGCCCGCATGGAACGAC	726
Db	661	AACCTCCCATCAGCCCCCATGAGACCGGTGCCTGTGAAGCTGAAGCCCGCATGGAACGAC	720
QY	727	CCCAAGGTGAAGCATGTGCGCTCTGACCGAGGAGGAAGATCAAGGCGCTGACCTGCACTTGC	786
Db	721	CCCAAGGTGAAGCATGTGCGCTCTGACCGAGGAGGAAGATCAAGGCGCTGACCTGCACTTGC	780
QY	787	GAGAGATGAGAGAGGAGGCAAGATCAACAAAGTGGGCGCCCGAGAAACCCCTTACAACAC	846
Db	781	GAGAGATGAGAGAGGAGGCAAGATCAACAAAGTGGGCGCCCGAGAAACCCCTTACAACAC	840
QY	847	CCCGTGTTCGCATCAAGAGAGAGACACCAAGTGCAGCAAGCTGTATGACTTTCGC	906
Db	841	CCCGTGTTCGCATCAAGAGAGAGACACCAAGTGCAGCAAGCTGTATGACTTTCGC	900
QY	907	GAGCTGAACAGCGCACCCAGGACTTCTGAGAGGTGCAGCTGGCATTCGCCACCCGCGC	966
Db	901	GAGCTGAACAGCGCACCCAGGACTTCTGAGAGGTGCAGCTGGCATTCGCCACCCGCGC	960
QY	967	GAGCTGAAGAGAGAGAGAGGTGACCGTGTGGAAGTGGGCGACGCTTACTTTCAGAGTG	1026
Db	961	GAGCTGAAGAGAGAGAGAGGTGACCGTGTGGAAGTGGGCGACGCTTACTTTCAGAGTG	1020
QY	1027	CCCGTGAACAGGACTTCCGCAAGTACACGCGCTTCAACATCCCGACGATCAACAGAG	1086
Db	1021	CCCGTGAACAGGACTTCCGCAAGTACACGCGCTTCAACATCCCGACGATCAACAGAG	1080
QY	1087	ACCCCGGCAATCGCTACAGTACCAAGTGTGCGCCAGAGGCTGGAAGGCGACGCCACGC	1146

```
Db 1081 ACCCCGGGATCCGCTACCAATGTAACAATGCTGCCCCAGGGGTGGAAGGGGACGCCCAAGC 1140
Qy 1147 ATCTTCCAGAGCAGATGACCAAGATCTGTAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCCAGAGCAGATGACCAAGATCTGTAGAGCCCTTCCGCGCCCGCAACCCCGAGATC 1200
Qy 1207 GTGATCTCA-----GGCCCCCTGTATCGTGGGCAAGGACCTGTAGATTCGGCCAGAC 1260
Db 1201 GTGATCTCAATGATGATGACCACTGTATCGTGGGCAAGGACCTGTAGATTCGGCCAGAC 1260
Qy 1261 CGCGCAAGATGAGGAGAGCTGTGCAAGACCTGTGCGTGGGGCTTCAACACCCCGAC 1320
Db 1261 CGCGCAAGATGAGGAGAGCTGTGCAAGACCTGTGCGTGGGGCTTCAACACCCCGAC 1320
Qy 1321 AAGAACACACAGAGAGAGCCCTTCTGTGTATGAGGCTACGAGCTGACCCCGACAG 1380
Db 1321 AAGAACACACAGAGAGAGCCCTTCTGTGTATGAGGCTACGAGCTGACCCCGACAG 1380
Qy 1381 TGGACCTGTGCAAGCCATCGAGCTGCGCGGCAAGAGAGAGCTGTGACCTGTGACCA 1440
Db 1381 TGGACCTGTGCAAGCCATCGAGCTGCGCGGCAAGAGAGAGCTGTGACCTGTGACCA 1440
Qy 1441 AAGCTGTGGGCAAGCTGAACTGGGCAAGCAGATCTACCCCGGATCAAGGTGGGCGAG 1500
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCAAGCAGATCTACCCCGGATCAAGGTGGGCGAG 1500
Qy 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCTTGAACGAGCATGTGCGCTTGAACGAGAG 1560
Db 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCTTGAACGAGCATGTGCGCTTGAACGAGAG 1560
Qy 1561 GCCGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGTGCGAGCCGTGCACGCGCTGAC 1620
Db 1561 GCCGAGCTGAGAGCTGCGCGCAAGACCGCGAGATCTGTGCGAGCCGTGCACGCGCTGAC 1620
Qy 1621 TACGACCCCGAGAGAGAGCTGTGCGCGAGATCTGAGAACGAGGCGGACGAGTGGAGC 1680
Db 1621 TACGACCCCGAGAGAGAGCTGTGCGCGAGATCTGAGAACGAGGCGGACGAGTGGAGC 1680
Qy 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGTGAGACCGGCAAGTACCCCAAGATCCG 1740
Db 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGTGAGAACCGGCAAGTACCCCAAGATCCG 1740
Qy 1741 ACCGCCCAACCAAGAGAGCTGAAAGCTGTGACCGAGCGGTGCAAGAAATCCGATGAG 1800
Db 1741 ACCGCCCAACCAAGAGAGCTGAAAGCTGTGACCGAGCGGTGCAAGAAATCCGATGAG 1800
Qy 1801 AAGCATCGATCTGTGGGCAAGACCCCAAGTTCGGCTCGCCCATTCAGAGAGAGCATGAG 1860
Db 1801 AAGCATCGATCTGTGGGCAAGACCCCAAGTTCGGCTCGCCCATTCAGAGAGAGCATGAG 1860
Qy 1861 GAGACCTGTGAGCCGACTACTGTGAGAGCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920
Db 1861 GAGACCTGTGAGCCGACTACTGTGAGAGCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920
Qy 1921 ACCCCCCCTGTGTGAAGCTGTGTATCAAGCTGTGAGAGAGAGCCCATTCGCGCGAG 1980
Db 1921 ACCCCCCCTGTGTGAAGCTGTGTATCAAGCTGTGAGAGAGAGCCCATTCGCGCGAG 1980
Qy 1981 ACCTTTCTAGTGAAGCGCGCCCGCAACCGGAGAACCAAGATCGGCAAGGCGCGCTACG 2040
Db 1981 ACCTTTCTAGTGAAGCGCGCGCGCAACCGGAGAACCAAGATCGGCAAGGCGCGCTACG 2040
Qy 2041 ACCGACCGGGGCGGAGAGATGTGTAGAGCTGTGACCGGACCAACCAAGAGAGAGCGAG 2100
Db 2041 ACCGACCGGGGCGGAGAGATGTGTAGAGCTGTGACCGGACCAACCAAGAGAGAGCGAG 2100
Qy 2101 CTGACAGGCAATTCAGCTGTGCGCTGTGAGAGACGGGAGCGAGGTGAACATCGTACGAC 2160
Db 2101 CTGACAGGCAATTCAGCTGTGCGCTGTGAGAGACGGGAGCGAGGTGAACATCGTACGAC 2160
Qy 2161 AGCGAATACGCTTGTGGGCAATTCAGAGCGGACCGGCAAGAGCGAGAGCGAGCTGTG 2220
Db 2161 AGCGAATACGCTTGTGGGCAATTCAGAGCGGACCGGCAAGAGCGAGAGCGAGCTGTG 2220
```

```
Qy 2221 AACCGATCATGAGCAGACTGATCAAGAGAGAGAGTGTACTGTAGAGTGTGCGCGC 2280
Db 2221 AACCGATCATGAGCAGACTGATCAAGAGAGAGAGTGTACTGTAGAGTGTGCGCGC 2280
Qy 2281 CACAGGGCATTCGGCGGCAAGAGAGATTCGACCAAGCTGTGTGAGAGAGGCAATCCGCAAG 2340
Db 2281 CACAGGGCATTCGGCGGCAAGAGAGATTCGACCAAGCTGTGTGAGAGAGGCAATCCGCAAG 2340
Qy 2341 GTGCTGTTCTGTGACGGGATGATGCGCGCATGTGTATCAAGTACATGACGACCTG 2400
Db 2341 GTGCTGTTCTGTGACGGGATGATGCGCGCATGTGTATCAAGTACATGACGACCTG 2400
Qy 2401 TACGTGGGCAAGCGGCGGCTTAGATGATTAAGCTTCCGCGGCTTAGACCCGGT 2457
Db 2401 TACGTGGGCAAGCGGCGGCTTAGATGATTAAGCTTCCGCGGCTTAGACCCGGT 2457

RESULT 10
ADCl3266
ID ADCl3266 standard; DNA; 2457 BP.
XX
XX
AC ADCl3266;
XX
XX
DT 18-DEC-2003 (first entry)
XX
XX
DE DNA of HIV construct p2Pol-opt_C SEQ ID NO 45.
XX
XX
KM expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
OS Human immunodeficiency virus.
XX
PN WO200304620-A2.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX PR 31-AUG-2001; 2001US-0316860P.
XX PR 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR) CHIRON CORP.
XX PA (UYST-) UNIV STELLENBOSCH.
XX
PI Zur Megeide J, Barnett SM, Lian Y, Engelbrecht S, Van Renenburg EJ,
XX
XX WPI; 2003-221593/21.
XX
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunisation, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 42; 301pp; English.
XX
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intranasally, intravenously,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
XX
XX
XX Sequence 2457 BP; 568 A; 830 C; 758 G; 301 T; 0 U; 0 Other;
XX
XX
XX Query Match 98.7%; Score 2430.2; DB 10; Length 2457;
XX Best Local Similarity 99.6%; Pred. No. 5.2e-293;
```

Matches 2448; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 7 GCCACCATGGCCGAGGCGCATGAGCCAGGCGACCAAGCGCCAAATCTGTGATGCAAGGCGACG 66
Db 1 GCCACCATGGCCGAGGCGCATGAGCCAGGCGACCAAGCGCCAAATCTGTGATGCAAGGCGACG 60
QY 67 AACTTCAGAGGGCCCCAGAGCGCATCATCATGTCTTCAACTGCGGCAAGAGAGGCGCCACATC 126
Db 61 AACTTCAGAGGGCCCCAGAGCGCATCATCATGTCTTCAACTGCGGCAAGAGAGGCGCCACATC 120
QY 127 GCCCGCACTGCGCGGCGCCCCCGCAAGAAAGGCTGTGAAAGTGTGCGCAAGAGGCGCAC 186
Db 121 GCCCGCACTGCGCGGCGCCCCCGCAAGAAAGGCTGTGAAAGTGTGCGCGCAAGAGGCGCAC 180
QY 187 CAGATGAAAGGACTGCAACCGAGGCGCAAGCCACTTCTTCGAGAGGAACTGTGCGCTTCCCG 246
Db 181 CAGATGAAAGGACTGCAACCGAGGCGCAAGCCACTTCTTCGAGAGGAACTGTGCGCTTCCCG 240
QY 247 CAGGCGCAAGGCGCGGAGTTCCCGAGCGAGCAAGAACCGGCGCCAGACGCCCAACAGCGCG 306
Db 241 CAGGCGCAAGGCGCGGAGTTCCCGAGCGAGCAAGAACCGGCGCCAGACGCCCAACAGCGCG 300
QY 307 GAGCTGCAAGTGTGCGGCGCAACCCCGCGAGCGAGGCTGTGCGCGGCGCGCGAC 366
Db 301 GAGCTGCAAGTGTGCGGCGCAACCCCGCGAGCGAGGCTGTGCGCGGCGCGCGAC 360
QY 367 CTGAACTTCCCGCAATCAACCTGTGTGAGCGCGCCCTGTGTGAGGACTCAAGGTGGCGCG 426
Db 361 CTGAACTTCCCGCAATCAACCTGTGTGAGCGCGCCCTGTGTGAGGACTCAAGGTGGCGCG 420
QY 427 CAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGTGTGTGAGAGATGAGC 486
Db 421 CAGATCAAGAGAGGCGCTGTGAGCAACCGGCGCGAGCAACCGTGTGTGAGAGATGAGC 480
QY 487 CTGCCCCGCAATGTGAAGCCCAAGATGATGTGCGCGCATGTGCGGCTTCACTCAAGTGTGCG 546
Db 481 CTGCCCCGCAATGTGAAGCCCAAGATGATGTGCGCGCATGTGCGGCTTCACTCAAGTGTGCG 540
QY 547 CAGTACGACACAGATCTGTATCGAGATCTGCGGCAAGAAAGCCATCGGACCGTGTGATC 606
Db 541 CAGTACGACACAGATCTGTATCGAGATCTGCGGCAAGAAAGCCATCGGACCGTGTGATC 600
QY 607 GGCAGCCAGCCCGGTGAACATCATTCGCGCGCAACATGTCTGACCCAGCTGTGACACCTTG 666
Db 601 GGCAGCCAGCCCGGTGAACATCATTCGCGCGCAACATGTCTGACCCAGCTGTGACACCTTG 660
QY 667 AACTTCCCATCAAGCCCATTCAGACCGTGTGCGGAGCTGAGAGCCCGGCAATGAGCGCG 726
Db 661 AACTTCCCATCAAGCCCATTCAGACCGTGTGCGGAGCTGAGAGCCCGGCAATGAGCGCG 720
QY 727 CCGAAGGTGAAGATGTGGCCCTGTGACCGAGAGAAAGATCAAGGCCCTGTGACCGGCAATGTG 786
Db 721 CCGAAGGTGAAGATGTGGCCCTGTGACCGAGAGAAAGATCAAGGCCCTGTGACCGGCAATGTG 780
QY 787 GAGGAGATGGAAGAGAGGCGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTTACAAAC 846
Db 781 GAGGAGATGGAAGAGAGGCGCAAGATCAACCAAGATCGGCCCGGAGAACCCCTTACAAAC 840
QY 847 CCGGTGTTCGCACTCAAGAAAGAGCAGACCAAGTGTGCGCAAGTGTGTGACTTTCGCG 906
Db 841 CCGGTGTTCGCACTCAAGAAAGAGCAGACCAAGTGTGCGCAAGTGTGTGACTTTCGCG 900
QY 907 GAGCTGAACAAGCGCAACCGAGACTTCTGTGAGGTGTGAGCTGTGGGCAATCCCGCACCGCGCG 966
Db 901 GAGCTGAACAAGCGCAACCGAGACTTCTGTGAGGTGTGAGCTGTGGGCAATCCCGCACCGCGCG 960
QY 967 GGCCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGTGCGAGCGCTTACTTCAAGCGTG 1026
Db 961 GGCCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGTGCGAGCGCTTACTTCAAGCGTG 1020
QY 1027 CCCCTGTGAGAGAGACTTTCGCAAGTACACCGGCTTTCACCATCCCGAGCATCAACAGAG 1086
Db 1021 CCCCTGTGAGAGAGACTTTCGCAAGTACACCGGCTTTCACCATCCCGAGCATCAACAGAG 1080

QY 1087 ACCCGGCGATCCGCTACCAAGTACAACTGTCTGCCCGACAGGCGTGAAGGCGACGCCAC 1146
Db 1081 ACCCGGCGATCCGCTACCAAGTACAACTGTCTGCCCGACAGGCGTGAAGGCGACGCCAC 1140
QY 1147 ATCTTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCCCGAGATC 1206
Db 1141 ATCTTCAGAGCAGCATGACCAAGATCTGTGAGCCCTTCCGCGCCGCAACCCCGAGATC 1200
QY 1207 GTGATCTACCA-----GGCCCCCTGTATCGTGTGGCAACGAACTGTGGAATGTGGCCAGCAC 1260
Db 1201 GTGATCTACCAAGTACATGAGCAACCTGTATCGTGTGGCAACGAACTGTGGAATGTGGCCAGCAC 1260
QY 1261 CGCGCCAGAGATCGAGAGACTGTGCGCAAGCACTGTGCGCTGTGGGCTTCAACACCCCGAC 1320
Db 1261 CGCGCCAGAGATCGAGAGACTGTGCGCAAGCACTGTGCGCTGTGGGCTTCAACACCCCGAC 1320
QY 1321 AAGAAACACAGAAAGAGCCCTCTTCTGTGTGATGTGGCTACGAGCTGCAACCCGACAAAG 1380
Db 1321 AAGAAACACAGAAAGAGCCCTCTTCTGTGTGATGTGGCTACGAGCTGCAACCCGACAAAG 1380
QY 1381 TGGACCTGTGAGCCCAATGAGCTGTGCGAGAAAGAGAACTGTGAACCTGTGAACATTCAG 1440
Db 1381 TGGACCTGTGAGCCCAATGAGCTGTGCGAGAAAGAGAACTGTGAACCTGTGAACATTCAG 1440
QY 1441 AAGCTGTGTGGCAAGCTGAACCTGTGCGACGACGATCTACCCCGGCAATCAAGGTGTGCGCAG 1500
Db 1441 AAGCTGTGTGGCAAGCTGAACCTGTGCGACGACGATCTACCCCGGCAATCAAGGTGTGCGCAG 1500
QY 1501 CTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGTGACCGACATGTGTCCCTGTGACCGAGAG 1560
Db 1501 CTGTGCAAGCTGTGCGGCGCGCAAGGCGCTGTGACCGACATGTGTCCCTGTGACCGAGAG 1560
QY 1561 GCGGAGCTGTGAGCTGTGGCGGAGAACCGGAGATCTGTGCGCGAGCCGTGTGACCGGCTGTAC 1620
Db 1561 GCGGAGCTGTGAGCTGTGGCGGAGAACCGGAGATCTGTGCGCGAGCCGTGTGACCGGCTGTAC 1620
QY 1621 TACGAGCCCGCAAGAGACTGTGTGCGCGAGATCTCAAGAGCAGGCGCAACGAGTGTGAC 1680
Db 1621 TACGAGCCCGCAAGAGACTGTGTGCGCGAGATCTCAAGAGCAGGCGCAACGAGTGTGAC 1680
QY 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAAAGCCGCGCAAGTACGCAAGATGCGC 1740
Db 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAAAGCCGCGCAAGTACGCAAGATGCGC 1740
QY 1741 ACCGCGCAACCAAGAGAGTGAAGAGCTGACCCGAGCCGTGCAAGAGATGTGCAATGTGAG 1800
Db 1741 ACCGCGCAACCAAGAGAGTGAAGAGCTGACCCGAGCCGTGCAAGAGATGTGCAATGTGAG 1800
QY 1801 AGCATGTGATCTGTGGGCAAGAGCCCGCAAGTTCCGCTGTGCGCATCAAGAGAGAGACTTGG 1860
Db 1801 AGCATGTGATCTGTGGGCAAGAGCCCGCAAGTTCCGCTGTGCGCATCAAGAGAGAGACTTGG 1860
QY 1861 GAGACCTGTGTGAGCCGACTTCTGTGAGGCGCAACCTTCCGCTGTGCGCATCAAGAGAGAGACTTGG 1920
Db 1861 GAGACCTGTGTGAGCCGACTTCTGTGAGGCGCAACCTTCCGCTGTGAGGAGTTCTGTGAAC 1920
QY 1921 ACCCGCGCGCTGTGTGAAGCTGTGTGTAACGAGCTGTGAGAAAGAGCCATCATTCGCGCGCGAG 1980
Db 1921 ACCCGCGCGCTGTGTGAAGCTGTGTGTAACGAGCTGTGAGAAAGAGCCATCATTCGCGCGCGAG 1980
QY 1981 ACCCTTCTAGTGTGAAGGCGCGCGCAACCGCGAGACCAAGATGTGGGCAAGGCGGCTACGTT 2040
Db 1981 ACCCTTCTAGTGTGAAGGCGCGCGCGCAACCGCGAGACCAAGATGTGGGCAAGGCGGCTACGTT 2040
QY 2041 ACCGACCGGCGCGCGCGCAAGATCTGTGAGCCCTGTGACCGAGACCAACCAAGAGACCGAG 2100
Db 2041 ACCGACCGGCGCGCGCGCGCAAGATCTGTGAGCCCTGTGACCGAGACCAACCAAGAGACCGAG 2100
QY 2101 CTGCAAGGCGCATTCAGCTGTGCTTGTGAGAGCAGCGGCAAGGAGTGAACATGTGTGACCGAC 2160
Db 2101 CTGCAAGGCGCATTCAGCTGTGCTTGTGAGAGCAGCGGCAAGGAGTGAACATGTGTGACCGAC 2160

CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polyepitides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 98.4%; Score 2422.6; DB 10; Length 2445;
Best Local Similarity 99.6%; Pred. No. 4.5e-292;
Matches 2441; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

QY 7 GGCACCATGCGCGAGCCCATGAGCCAGGCGCAAGCCAGGCGCAATCTCTGATGCAAGGCGCAGC 66
DB 1 GCCACCATGCGCGAGCCCATGAGCCAGGCGCAAGCCAGGCGCAATCTCTGATGCAAGGCGCAGC 60
QY 67 AACTTCAGAGGCGCCCAAGGCGCATCATCAAGTCTTCACTGCGCGCAAGAGAGGCGCAATC 126
DB 61 AACTTCAGAGGCGCCCAAGGCGCATCATCAAGTCTTCACTGCGCGCAAGAGAGGCGCAATC 120
QY 127 GCCCGCACTGCG 186
DB 121 GCCCGCACTGCG 180
QY 187 CAGATGAAGGACTGCAACCGAGGCG 246
DB 181 CAGATGAAGGACTGCAACCGAGGCG 240
QY 247 CAGGCGCAAGGCG 306
DB 241 CAGGCGCAAGGCG 300
QY 307 GAGCTGCAAGTGTGCG 366
DB 301 GAGCTGCAAGTGTGCG 360
QY 367 CTGGAATTTCCCCCGAAGTCAACCTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
DB 361 CTGGAATTTCCCCCGAAGTCAACCTCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
QY 427 CAGATCAAGAGGCG 486
DB 421 CAGATCAAGAGGCG 480
QY 487 CTGCGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATGCGCGCGCTTTCATCAAGGTGCGC 546
DB 481 CTGCGCGCGCAAGTGAAGCCCAAGATGATGCGCGCGCATGCGCGCGCTTTCATCAAGGTGCGC 540
QY 547 CAGTACGACCAAGTCTCTGATCGAGTCTGCGCGCAAGAGGCGCATCGGCACTGTGCTGATC 600
DB 541 CAGTACGACCAAGTCTCTGATCGAGTCTGCGCGCAAGAGGCGCATCGGCACTGTGCTGATC 600
QY 607 GGGCCCAACCCCGTGAACATCATCGGCGCGCAATGCTGAAAGCTGAAGCCCGGCAATGAAAGGCG 666
DB 601 GGGCCCAACCCCGTGAACATCATCGGCGCGCAATGCTGAAAGCTGAAGCCCGGCAATGAAAGGCG 660
QY 667 AACTTCCCATCAGCCCATCAGACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
DB 661 AACTTCCCATCAGCCCATCAGACCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
QY 727 CCCAAGGTGAAGAGTGGCG 786
DB 721 CCCAAGGTGAAGAGTGGCG 780
QY 787 GAGGAGATGAGAGAGAGGCGCAAGATCAGCAAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 846
DB 781 GAGGAGATGAGAGAGAGGCGCAAGATCAGCAAGATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 847 CCGGTGTTCCCATCAAG 906
DB 841 CCGGTGTTCCCATCAAG 900
QY 907 GAGGTGAACAGCGCGCACTTCTGTGAGAGTGAAGTGGGCGCATCCCGCGCGCGCGCGCGCGCGCG 966

DB 901 GAGGTGAACAGCGCGCACTTCTGTGAGAGTGAAGTGGGCGCATCCCGCGCGCGCGCGCGCGCG 960
QY 967 GGCCTGAAG 1026
DB 961 GGCCTGAAG 1020
QY 1027 GGCCTGAAG 1086
DB 1021 GGCCTGAAG 1080
QY 1087 ACCCGCGCATCGCTACCAAGTCAACAGTGTCTCCCAAGGCGTGAAGAGGCGAGCCCGCAGC 1146
DB 1081 ACCCGCGCATCGCTACCAAGTCAACAGTGTCTCCCAAGGCGTGAAGAGGCGAGCCCGCAGC 1140
QY 1147 ATCTTCAG 1206
DB 1141 ATCTTCAG 1200
QY 1207 GTGATCTACAGGCG 1266
DB 1201 GTGATCTACAGGCG 1260
QY 1267 AAGATCGAG 1326
DB 1261 AAGATCGAG 1320
QY 1327 CACAG 1386
DB 1321 CACAG 1374
QY 1387 GTGCGAGCGCGAG 1446
DB 1375 GTGCGAGCGCGAG 1434
QY 1447 GTGCGAG 1506
DB 1435 GTGCGAG 1494
QY 1507 AAGCTGCTGCG 1566
DB 1495 AAGCTGCTGCG 1554
QY 1567 CTGGAAGTGTGCGCGAG 1626
DB 1555 CTGGAAGTGTGCGCGAG 1614
QY 1627 CCCAGCAAG 1686
DB 1615 CCCAGCAAG 1674
QY 1687 ATCTACAG 1746
DB 1675 ATCTACAG 1734
QY 1747 CAGACCAAG 1806
DB 1735 CAGACCAAG 1794
QY 1807 GTGATCTGAGGAG 1866
DB 1795 GTGATCTGAGGAG 1854
QY 1867 TGGTGAACGAGCTTACGAGGCGAGCTGAGATCCCGAGTGGAGGTTCTGAAACACCCCGC 1926
DB 1855 TGGTGAACGAGCTTACGAGGCGAGCTGAGATCCCGAGTGGAGGTTCTGAAACACCCCGC 1914
QY 1927 CCGCTGTGAAGAGTGTGTGATACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
DB 1915 CCGCTGTGAAGAGTGTGTGATACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1974
QY 1987 TACGTGAACGAGCGCGCGAG 2046
DB 1975 TACGTGAACGAGCGCGCGAG 2034

QY 914 ACAAGCGACCCAGAGCTTGTGGAGGTGACGTGGGCAATCCGCCACCCCGCGGCTGA 973
DB 2387 ACAAGGCGACCCAGAGCTTGTGGAGGTGACGTGGGCAATCCGCCACCCCGCGGCTGA 2446
QY 974 AGAAGAAAGAGCGGAGCGGTGCTGGAGGTGGGCGACCGCTTACGCGTGGCCCTGG 1033
DB 2447 AGAAGAAAGAGCGGAGCGGTGCTGGAGGTGGGCGACCGCTTACGCGTGGCCCTGG 2506
QY 1034 ACAGAGCTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAGAGACCCCG 1093
DB 2507 ACAGAGCTTCCGCAAGTACACCGCTTCAACATCCCGAGCATCAACAAGAGACCCCG 2566
QY 1094 GCATCCGCTACCAAGTCAACGTGCTCCCGAGGCTGGAGAGGCGACCCCGAGCATCTTC 1153
DB 2567 GCATCCGCTACCAAGTCAACGTGCTCCCGAGGCTGGAGAGGCGACCCCGAGCATCTTC 2626
QY 1154 AGAGCGAGTACCAAGATCTGGAGGCGCTTCCGCGCGCGCAACCCCGAGATCTGTATCT 1213
DB 2627 AGAGCGAGTACCAAGATCTGGAGGCGCTTCCGCGCGCGCAACCCCGAGATCTGTATCT 2686
QY 1214 ACCAGGCGCGCTGTAGTGGGCGAGCACTTGGAGTGGGCGACCGCGCCAGATG 1273
DB 2687 ACCAGGCGCGCTGTAGTGGGCGAGCACTTGGAGTGGGCGACCGCGCCAGATG 2746
QY 1274 AGAGCTGGCGCAAGCACTTGTGCGTGGGCTTCAACACCCCGCAAGAGACACAGA 1333
DB 2747 AGAGCTGGCGCAAGCACTTGTGCGTGGGCTTCAACACCCCGCAAGAGACACAGA 2806
QY 1334 AGAGGCGCGCTTCTGTGTGATGGGCTAGAGCTGACCCCGCAAGTGAACCTGTGACG 1393
DB 2807 AGAGGCGCGCTTCTGTGTGATGGGCTAGAGCTGACCCCGCAAGTGAACCTGTGACG 2860
QY 1394 CCATGAGCTGCCCGAAGAGAGAGTGAACGTGAACGATCCAGAGAGCTGTGGGCA 1453
DB 2861 CCATGAGCTGCCCGAAGAGAGAGTGAACGTGAACGATCCAGAGAGCTGTGGGCA 2920
QY 1454 AGCTGAATGGGCGACCGAGATCTACCCCGGCAATCAAGGTGGCGCAGCTGTGCAAGCTGG 1513
DB 2921 AGCTGAATGGGCGACCGAGATCTACCCCGGCAATCAAGGTGGCGCAGCTGTGCAAGCTGG 2980
QY 1514 TGGCGCGCGCGCAAGGCGCTTGAACGATGTGCGCTTGAACCGAGAGGCGGAGCTGAGC 1573
DB 2981 TGGCGCGCGCGCAAGGCGCTTGAACGATGTGCGCTTGAACCGAGAGGCGGAGCTGAGC 3040
QY 1574 TGGCGAGAACCGCGAGATCTTGGCGGAGCGGTGACCGCGGCTTGAACCGAGCA 1633
DB 3041 TGGCGAGAACCGCGAGATCTTGGCGGAGCGGTGACCGCGGCTTGAACCGAGCA 3100
QY 1634 AGGACCTGTGGCGGAGATCTCAAGAGAGGCGCAACGAGTGAACCTTACAGATCTACC 1693
DB 3101 AGGACCTGTGGCGGAGATCTCAAGAGAGGCGCAACGAGTGAACCTTACAGATCTACC 3160
QY 1694 AGGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCACACA 1753
DB 3161 AGGAGCGCTTCAAGAACTTGAAGACCGGCAAGTACCGCAAGATGCGCACCGCCACACA 3220
QY 1754 AGGACGTGAAGAGCTGACCGGAGCGGTGCAAGAGATGCGCAAGAGATGTATCT 1813
DB 3221 AGGACGTGAAGAGCTGACCGGAGCGGTGCAAGAGATGCGCAAGAGATGTATCT 3280
QY 1814 GGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGACTGGGAGACTGTGGGA 1873
DB 3281 GGGGCAAGACCCCAAGTTCCGCTGCGCATCCAGAGAGAGACTGGGAGACTGTGGGA 3340
QY 1874 CCGACTACTGGCGAGGCGCACTTGAATCCCGAGTGGAGATTCTGGAACACCCCGCTGG 1933
DB 3341 CCGACTACTGGCGAGGCGCACTTGAATCCCGAGTGGAGATTCTGGAACACCCCGCTGG 3400
QY 1934 TGAAGCTGTGATCAGCTGAGAGAGGCGCATCATCGGCGCGAGACTTCTAGCTGG 1993
DB 3401 TGAAGCTGTGATCAGCTGAGAGAGGCGCATCATCGGCGCGAGACTTCTAGCTGG 3460

QY 1994 AGGCGCGCGCAACCGCGAGACCAAGATCGGAGAGCGCGCTTACGAGACCGCGGCGC 2053
DB 3461 AGGCGCGCGCAACCGCGAGACCAAGATCGGAGAGCGCGCTTACGAGACCGCGGCGC 3520
QY 2054 GGCAGAGATCGTGAAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATC 2113
DB 3521 GGCAGAGATCGTGAAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGACGAGCCATC 3580
QY 2114 AGCTGCGCTTGAAGAGACGCGGACGAGTGAACATCTGTACCGACAGCTTACGCCCC 2173
DB 3581 AGCTGCGCTTGAAGAGACGCGGACGAGTGAACATCTGTACCGACAGCTTACGCCCC 3640
QY 2174 TGGGCAATCAAGGCGCGCGCGCGCAAGAGAGAGCGAGAGAGAGAGAGAGAGAGAG 2233
DB 3641 TGGGCAATCAAGGCGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3700
QY 2234 AGCAGCTGATCAAG 2293
DB 3701 AGCAGCTGATCAAG 3760
QY 2294 GCGGCAAGAGAGAGATCGAACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
DB 3761 GCGGCAAGAGAGAGATCGAACAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3820
QY 2354 AGGCAATCATGCGCGCGCATCTGTATCTACAGTACATGAGACGACTGTACGTGGGAGG 2413
DB 3821 AGGCAATCATGCGCGCGCATCTGTATCTACAGTACATGAGACGACTGTACGTGGGAGG 3880
QY 2414 GGGGCGCTTGAAGATTAAGAGCTTCCGCGGCGCTAGACAGCGGT 2457
DB 3881 GGGGCGCTTGAAGATTAAGAGCTTCCGCGGCGCTAGACAGCGGT 3924

RESULT 14
ADCl3231
ID ADCl3231 standard; DNA; 3930 BP.
XX
AC ADCl3231;
XX
AC 18-DEC-2003 (first entry)
XX
DE DNA of HIV construct GagComp1PolmutAtc_C SEQ ID NO 10.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
KW Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
OS Human immunodeficiency virus.
XX
PN WO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PP 05-UTL-2002; 2002WO-US021420.
XX
PR 05-UTL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHTR) CHIRON CORP.
PA (OYST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Llan Y, Engelbrecht S, Van Rensburg EJ;
XX
XX WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure; Fig 7; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a

CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15^{cas}, p17, Tat, Pro, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intranasally, intravenously,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunization, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.

XX Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;

Query Match 98.0%; Score 2414; DB 10; Length 3930;

Best Local Similarity 99.5%; Pred. No. 4,9e-291;
Matches 2433; Conservative 0; Mismatches 5; Indels 6; Gaps 1;

```
QY 14 TGGCGAGGCGCATGAGCGGCGCACCAGCGCAATCTGTATGACGCGCAACTTCA 73
DB 1487 TGGCGAGGCGCATGAGCGGCGCACCAGCGCAATCTGTATGAGCGCGCAACTTCA 1546
QY 74 AGGGCGCCGAGGCGCATGAGCGGCGCACCAGCGCAATCTGTATGAGCGCGCAACTTCA 133
DB 1547 AGGGCGCCGAGGCGCATGAGCGGCGCACCAGCGCAATCTGTATGAGCGCGCAACTTCA 1606
QY 134 ACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAAAGTGCAGAGAGGCGCAACAGATGA 193
DB 1607 ACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAAAGTGCAGAGAGGCGCAACAGATGA 1666
QY 194 AGGACTGACCGAGCGCGCGCAACCTTCTTCGCGAGGACCTTGCTTCCCGCAGGCA 253
DB 1667 AGGACTGACCGAGCGCGCGCAACCTTCTTCGCGAGGACCTTGCTTCCCGCAGGCA 1726
QY 254 AGGGCGCGGAGTTTCCCGAGCGAGAGAACCGCGCGCAACGCGCGCGCGCGCGCTGC 313
DB 1727 AGGGCGCGGAGTTTCCCGAGCGAGAGAACCGCGCGCAACGCGCGCGCGCGCGCTGC 1786
QY 314 AGGTGCGCGCGCGCAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 373
DB 1787 AGGTGCGCGCGCGCAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 1846
QY 374 TCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 433
DB 1847 TCCCGCGAGTCAACCTGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 1906
QY 434 AGGAGCGCGCGCGCGCAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 493
DB 1907 AGGAGCGCGCGCGCGCAACCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 1966
QY 494 GCAAGTGAAGCGCGCAAGATGATCGGCGCGCATCGCGCGCTTCAATCAAGTGCAGAGTAG 553
DB 1967 GCAAGTGAAGCGCGCAAGATGATCGGCGCGCATCGCGCGCTTCAATCAAGTGCAGAGTAG 2026
QY 554 ACCAGATCTGTATGAGATCTGCGGCAAGAGCGCATTCGCGACCGTGTGATCGCGCGCA 613
DB 2027 ACCAGATCTGTATGAGATCTGCGGCAAGAGCGCATTCGCGACCGTGTGATCGCGCGCA 2086
QY 614 CCCCCGTGAACATTCGCGCGCGCGCAACATGCTGACCGAGTGTGCGTGCACCTTGAATTCC 673
DB 2087 CCCCCGTGAACATTCGCGCGCGCGCAACATGCTGACCGAGTGTGCGTGCACCTTGAATTCC 2146
QY 674 CCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 733
DB 2147 CCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 2206
QY 734 TGAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 793
DB 2207 TGAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 2266
QY 794 TGAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 853
DB 2267 TGAAGCAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 2326
```

```
QY 854 TGGCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 913
DB 2327 TGGCGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2386
QY 914 ACAAGCGCACCGAGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 973
DB 2387 ACAAGCGCACCGAGAGCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2446
QY 974 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
DB 2447 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
QY 1034 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1093
DB 2507 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2566
QY 1094 GCATCTGCTACCAAGTACAAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 1153
DB 2567 GCATCTGCTACCAAGTACAAAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGC 2626
QY 1154 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1213
DB 2627 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2686
QY 1214 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1273
DB 2687 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2746
QY 1274 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1333
DB 2747 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2806
QY 1334 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1393
DB 2807 AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2860
QY 1394 CCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1453
DB 2861 CCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2920
QY 1454 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1513
DB 2921 AGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2980
QY 1514 TGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1573
DB 2981 TGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3040
QY 1574 TGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1633
DB 3041 TGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3100
QY 1634 AGGAGCTGTGCGCGCGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1693
DB 3101 AGGAGCTGTGCGCGCGAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3160
QY 1694 AGGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1753
DB 3161 AGGAGCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3220
QY 1754 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1813
DB 3221 ACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3280
QY 1814 GGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1873
DB 3281 GGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3340
QY 1874 CCGAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1933
DB 3341 CCGAGCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3400
QY 1934 TGAAGCTGTGTATCAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1993
```


Db 2267 TGGAGAGAGGAGGAGGATCAACAGATGGGCCCCGAGAACCCCTTACAAACACCCCGGTGT 2326
QY 864 TGGCATCAAGAAAGAAAGGACAGACCAAGTGGCGCAAGCTGTGTGACTTCCCGAGCTGA 913
Db 2327 TGGCATCAAGAAAGAAAGGACAGACCAAGTGGCGCAAGCTGTGTGACTTCCCGAGCTGA 2386
QY 914 ACAAGCGACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCGCCACCCCGCGGCTGA 973
Db 2387 ACAAGCGACCCAGGACTTCTGGAGGTGCAGCTGGGCATCCGCCACCCCGCGGCTGA 2446
QY 974 AGAAGAGAAAGAGCTGACCTGTGCTGTGAAGTGGGCGACGCTTACCTTCAAGCTGCCCTGG 1033
Db 2447 AGAAGAGAAAGAGCTGACCTGTGCTGTGAAGTGGGCGACGCTTACCTTCAAGCTGCCCTGG 2506
QY 1034 AGAAGAGCTTCCGCAAGTACACGCGCTTACACATCCCGAGCATCAACAAACGAGACCCCG 1093
Db 2507 AGAAGAGCTTCCGCAAGTACACGCGCTTACACATCCCGAGCATCAACAAACGAGACCCCG 2566
QY 1094 GCATCCGCTACAGTACAAAGTGTGCTGCCAGAGGCTGAAAGGCGAGCCAGCATCTTCC 1153
Db 2567 GCATCCGCTACAGTACAAAGTGTGCTGCCAGAGGCTGAAAGGCGAGCCAGCATCTTCC 2626
QY 1154 AGAGCAGCATGACCAAGATCTGTGAAGCCTTTCGCGCGCGCAACCCCGAGATCGTATCT 1213
Db 2627 AGAGCAGCATGACCAAGATCTGTGAAGCCTTTCGCGCGCGCAACCCCGAGATCGTATCT 2686
QY 1214 ACCAAGCCCCCTGTACGTGTGGCAGCGACCTGTGAAGTGGGCGAGACCGCGCAAGATCG 1273
Db 2687 ACCAAGCCCCCTGTACGTGTGGCAGCGACCTGTGAAGTGGGCGAGACCGCGCAAGATCG 2746
QY 1274 AGAGCTGGCAGAGACCTGTGCGCTGGGGCTTCAACACCCCGCAAGAGACAGCA 1333
Db 2747 AGAGCTGGCAGAGACCTGTGCGCTGGGGCTTCAACACCCCGCAAGAGACAGCA 2806
QY 1334 AGAGCCCCCTTCTGTGTATGGGCTACGAGCTGCAACCCCGCAAGTGAACCTGTGAC 1393
Db 2807 AGAGCCCCCTTCTGTGTATGGGCTACGAGCTGCAACCCCGCAAGTGAACCTGTGAC 2860
QY 1394 CCATGAGGCTGGCCGAGAGAGAGAGTGTGACCGTGAACATCCAGAAAGCTGTGGGCA 1453
Db 2861 CCATGAGGCTGGCCGAGAGAGAGAGTGTGACCGTGAACATCCAGAAAGCTGTGGGCA 2920
QY 1454 AGCTGAAGTGGGCGAGCGAGATCTACCCCGGATCAAGTGGCGGAGCTGTGCAAGCTGC 1513
Db 2921 AGCTGAAGTGGGCGAGCGAGATCTACCCCGGATCAAGTGGCGGAGCTGTGCAAGCTGC 2980
QY 1514 TGGCGGCGCGCAAGGCTTGACCGACATGTGCCCCCTGAACGAGAGGCGGAGCTGAGC 1573
Db 2981 TGGCGGCGCGCAAGGCTTGACCGACATGTGCCCCCTGAACGAGAGGCGGAGCTGAGC 3040
QY 1574 TGGCGGAGAACCGCGAGATCTCTGCGCGGAGCCCGTGCACGCGGTGTACTACACCCAGCA 1633
Db 3041 TGGCGGAGAACCGCGAGATCTCTGCGCGGAGCCCGTGCACGCGGTGTACTACACCCAGCA 3100
QY 1634 AGGACTGTGGCCGAGATCCAGAAAGAGGCGACAGACAGTGAACCTTACAGATCTTACC 1693
Db 3101 AGGACTGTGGCCGAGATCCAGAAAGAGGCGCGACAGTGAACCTTACAGATCTTACC 3160
QY 1694 AGGAGCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGTGGCACCGGCCACACCA 1753
Db 3161 AGGAGCCTTCAAGAACCTGAAGACCGGCAAGTACGCGCAAGTGGCACCGGCCACACCA 3220
QY 1754 ACAGCTGAAGACAGTGAACCGAGGCGGTGCAGAAAGATCGCATGTGAGAGATCTGTATCT 1813
Db 3221 ACAGCTGAAGACAGTGAACCGAGGCGGTGCAGAAAGATCGCATGTGAGAGATCTGTATCT 3280
QY 1814 GGGGCAAGACCCCAAGTTCCGCTGGCCCATCCAGAAAGAGACTGTGGAGACTGTGTGA 1873
Db 3281 GGGGCAAGACCCCAAGTTCCGCTGGCCCATCCAGAAAGAGACTGTGGAGACTGTGTGA 3340
QY 1874 CCGACTTACTGGAGGCGCACTGTGATCCCGAGTGGAGTTGTGTGAACACCCCGCCCTGG 1933

Db 3341 CCGACTTACTGGAGGCGCACTGTGATCCCGAGTGGAGTTGTGTGAACACCCCGCCCTGG 3400
QY 1934 TGAAGCTGTGTATCCAGCTGAGAGAGGCCATCATGTGGGCGCGAGACCTTCTTACGTGG 1993
Db 3401 TGAAGCTGTGTATCCAGCTGAGAGAGGCCATCATGTGGGCGCGAGACCTTCTTACGTGG 3460
QY 1994 ACGGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGAACCGACCGGGCC 2053
Db 3461 ACGGCGCGCGCAACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGAACCGACCGGGCC 3520
QY 2054 GCGAGAGATGTGTAGCTTGAACCGAGACCAACCAAGAGACCGAGCTTGCAGGCTATCC 2113
Db 3521 GCGAGAGATGTGTAGCTTGAACCGAGACCAACCAAGAGACCGAGCTTGCAGGCTATCC 3580
QY 2114 AGCTGGCCCTTGCAGGACAGCGCGAGCGGTGAACATGTGTACCGACAGCGCATGCGCC 2173
Db 3581 AGCTGGCCCTTGCAGGACAGCGCGAGCGGTGAACATGTGTACCGACAGCGCATGCGCC 3640
QY 2174 TGGGATCATCCAGGCGCGAGCCCGACAAAGCGAGAGCGAGCTGTGAACAGATCATCG 2233
Db 3641 TGGGATCATCCAGGCGCGAGCCCGACAAAGCGAGAGCGAGCTGTGAACAGATCATCG 3700
QY 2234 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGTGGTGCCTGCCACAAAGGCGATCG 2293
Db 3701 AGCAGCTGATCAAGAGAGAGAGGTGTACTGTAGCTGTGGTGCCTGCCACAAAGGCGATCG 3760
QY 2294 GCGGCAAGACAGATCGACAGCTGTGTGACAGGAGCATCCGCAAGGTGCTTTCCTGG 2353
Db 3761 GCGGCAAGACAGATCGACAGCTGTGTGACAGGAGCATCCGCAAGGTGCTTTCCTGG 3820
QY 2354 ACGGATGATGTGGCGGATCTGTGTACTACAGTACATGAGACGACTGTACGTGGCAGCG 2413
Db 3821 ACGGATGATGTGGCGGATCTGTGTACTACAGTACATGAGACGACTGTACGTGGCAGCG 3880
QY 2414 GCGGCTTAAAGATGATTTAAAGCTTCCCGGGGCTTAAAGCATCGGT 2457
Db 3881 GCGGCTTAAAGATGATTTAAAGCTTCCCGGGGCTTAAAGCATCGGT 3924

Search completed: December 30, 2005, 08:57:01
Job time : 1304.67 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8697.67 Seconds
(without alignments) 13249.138 Million cell updates/sec

Title: US-09-610-313B-31

Perfect score: 2463

Sequence: 1 gtcgacgccaccatgcccga.....gggctagcaccggcgaattc 2463

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_hic:*
5: gb_esc4:*
6: gb_esc5:*
7: gb_esc6:*
8: gb_esc7:*
9: gb_g881:*
10: gb_g882:*
11: gb_g883:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136.4	5.5	330	10	CL293849
2	98.4	4.0	2886	10	CL967755 OaIFCC015
3	90.2	3.7	1781	8	CK099623 RRCm0260
4	86.8	3.5	2598	4	AY103647 Zea mayb
5	86.6	3.5	1398	10	CL2647380 AIAA-aaf3
6	86	3.5	743	10	CZ247380 AIAA-aaf3
7	85.4	3.5	951	3	BM321451 rockefell
8	85.4	3.5	2031	10	CL974989 OaIFCC042
9	84	3.4	1725	10	CL978463 OaIFCC031
10	80.8	3.3	1060	10	CW922203 BDCAR297R
11	80.8	3.3	3069	10	CL973391 OaIFCC025
12	80.6	3.3	869	7	CK159167 FGS04056
13	80.6	3.3	892	10	CZ216254 AIAA-aaf2
14	80.6	3.3	1941	10	CL971508 OaIFCC021
15	80.2	3.3	1132	3	BM320864 rockefell
16	79.6	3.2	1165	3	BM320900 rockefell
17	79.6	3.2	11691	10	CL962901 OaIFCC008
18	79	3.2	867	3	BM321430 rockefell
19	78.6	3.2	2682	10	CL969033 OaIFCC017
20	78.4	3.2	873	7	CO443531 MGCCL1006
21	78.2	3.2	2853	10	CL974397 OaIFCC025
22	78	3.2	1485	10	CL970981 OaIFCC020

23	78	3.2	2313	10	CL982362	CL982362 OaIFSC047
24	78	3.2	2454	10	CL975440	CL975440 OaIFCC027
25	77.4	3.1	1509	10	CL958255	CL958255 OaIFCC002
26	77.4	3.1	2559	10	CL982027	CL982027 OaIFSC046
27	76.6	3.1	1550	3	BM321022	BM321022 rockefell
28	76.4	3.1	2028	10	CL979437	CL979437 OaIFCC033
29	76.2	3.1	1401	10	CL962721	CL962721 OaIFCC038
30	76.2	3.1	1680	10	CL982770	CL982770 OaIFSC049
31	76	3.1	2793	10	CL948321	CL948321 OaIFSC004
32	75	3.0	862	3	BM321023	BM321023 rockefell
33	75	3.0	1734	10	CL948553	CL948553 OaIFSC005
34	74.6	3.0	1410	10	CL947314	CL947314 OaIFSC004
35	74.6	3.0	1989	10	CL980973	CL980973 OaIFSC035
36	74.4	3.0	914	9	BZ568300	BZ568300 pac82-164
37	74	3.0	1071	10	CW922105	CW922105 BDCB195TF
38	74	3.0	1290	10	CL972679	CL972679 OaIFCC023
39	74	3.0	1386	11	DQ045165	DQ045165 Homo sapi
40	74	3.0	2072	4	CR603312	CR603312 full-length
41	73.2	3.0	853	3	BM321393	BM321393 rockefell
42	73	3.0	2151	10	CL972100	CL972100 OaIFCC041
43	72.8	3.0	566	3	BM587428	BM587428 170006873
44	72.6	2.9	788	6	CB643171	CB643171 OSUNB03L
45	72.4	2.9	753	9	CC675888	CC675888 OGMCO51TH

ALIGNMENTS

RESULT 1
LOCUS CL293849/c 330 bp DNA linear GSS 12-FEB-2004
DEFINITION 02S0349-08A1-C03 UniformMu MuTARIL Library Zea mays genomic clone
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 330)
REFERENCE Lathaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
AUTHORS Sequence tagged transposon insertions from the UniformMu maize
TITLE population
JOURNAL Unpublished (2003)
COMMENT Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmc@ufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
02S0349-08, Primer set: A
Class: transposon insertion site.
Location/Qualifiers
1..330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m9)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="02S0349-08A1-C03"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon
insertions in Mu inactive lines were extracted from the
UniformMu maize population by the thermo asymmetric
interlaced PCR (TAIL) protocol using primers specific for
the Mu terminal inverted repeat and a set of 16 arbitrary
primers. Amplicons were size enriched using Sephadex 400
spin columns and cloned into the TOPO PCR4 vector."

FEATURES

source

ORIGIN

Query Match	5.5%	Score 136.4	DB 10	Length 330
Best Local Similarity	66.4%	Pred. No. 2.7e-14		
Matches 211	Conservative 0	Mismatches 106	Indels 1	Gaps 1
QY	421	GGCGGCGCAGATCAAGAGGCGCCCTGTGTGACACCGGCGCGACGACCGGTGTGGAGGAG	480	
Db	329	GGGGGGGCGAGCTGGAGAGAGCTTATTAGATACAGAGCGAGATGATACAGTATTAGAGAA	270	
QY	481	ATGAGCCCTGGCCCGGCAAGTGGAGGCCCAAGATGATCGGCGGCGATCGGCGGCTTCATCAAG	540	
Db	269	ATGATTTTGGACAGAGAAAGATGAGAAACCAAAAATGATAGGGGGGAAATGGAGTTTATCAAA	210	
QY	541	GTGGCGCCAGTACAGACCAATCTGTATCGAGATCTGCGGACAAAGAGCCATCGGACCGTG	600	
Db	209	GTAAACACGTATGATTCAGGTACCCATAGAAATCTGTGGGCATTAAGCTATAGGTACGGTA	150	
QY	601	CTGATCGGCGCCCAACCCCGTGAACATCATCGGCGCGAATGATCTGACCCAGCTGGAGCTGC	660	
Db	149	TTAGTAGAGACCTACACCTGTCAACATTAATTGGAGAAATCTTTGACTCAGATTGG-TCG	91	
QY	661	ACCCTGAACCTTCCCATCATCAGCCCATCGACACCGTGCCCGTGAAGCTGAAGCCCGGCAATG	720	
Db	90	ACCTTAATTTTCCCATTAGTCTCATTTAGTAACCTGTACCACTAATTAAGCAGAGAAATG	31	
QY	721	GACGCGCCCAAGGTGAAG	738	
Db	30	GATGGCCCAAAAGTAAGG	13	

RESULT 2	2886 bp	DNA	linear	GSS 21-SEP-2000
LOCUS				
DEFINITION				
	CL967755			
	OSJFCC015178	Oryza sativa	Express Library Oryza sativa (Indica	
		cultivar-group)	genomic, genomic survey sequence.	

ACCESSION	CL967755
VERSION	CL967755.1
GI	52390149

KEYWORDS
SOURCE
ORGANISM
GSS.
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)
Oryza sativa (indica cultivar-group)

REFERENCE

1 (bases 1 to 286)

Ehretiacidae; Oryzeae; Oryza.

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Kuararctia; Viridiplantae; streptophyta; Embryophytes; Tracheophyta;

Kuararctia; Viridiplantae; streptophyta; Embryophytes; Tracheophyta;

<p>AUTHORS</p> <p>Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W., and Wang, J.</p>	<p>TITLE</p> <p>An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis</p>	<p>JOURNAL</p> <p>Unpublished (2004)</p>	<p>COMMENT</p> <p>Contact: Chen Chen</p>
--	---	---	---

Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

FEATURES	SOURCE
Location/Qualifiers	
1..2886	
/organism="Oryza sativa (indica cultivar-group)"	
/mol_type="genomic DNA"	
/db_xref="taxon:39946"	
/clone_lib="Oryza sativa Express Library"	
/note="Oryza sativa exon trapped genomic sequences"	
ORIGIN	

Query Match 4.0%; Score 98.4; DB 10; Length 2886;
Best local similarity 43.2%; Pred No. 2.1e-07;
Matches 709; Conservative 0; Mismatches 906; Indels 27; Gaps 4

189 GATGAGGACTGCACCCGAGCGGACGACCACTTCTCCGAGAGACTGCGCTTCCCA 248

Db	219	GGTATGAGAGCGCCGCCCATCATGCGCATCTGCTCTGCGCAACGGACGGACGGACCGCC	278
OY	249	GGGCAAGAGCCCGCAGATTTCCTCCAGCGAGAGAACTGGCCCAACAGCCCAACGACCGCGA	308
Db	279	CGACTGGACAGACTTCGTCCGCGCATGCTGCTCTCTGTCATCACTCCACATCAAGTTT	338
OY	309	GCTGCAGATGTCG---GGCGACAACTCCCGCAGAGAGCGCGGCGCGACGGCCAGGC	365
Db	339	CATGAGAGAGAACATGCTCCGGGAAACGCGCGCGGCGCTCATGCGCGGCTCTCGGCCAA	398
OY	366	CCTGAACCTTCCCGAGATCACTCTGAGGACGCGCCCTGGTGTGACATCAAGTGGGCGG	425
Db	399	GACCAAGGTGCTGAGGACGGGAAATGGCAGAGACAGAGCGCTTCATCTCTGTTCCCG	458
OY	426	CCAGATCAAGAGAGGCGCTGTGACACCGGCGCGACGACACCGTGTGAGAGATGAG	485
Db	459	CGACATCATGACATCAAGGCTCGGCGCATCATCCCGCGACCGCGGCGCTGTCAAGGG	518
OY	486	CTTGCCTCGGCGAATGAGAAAGCCCAAGATGATGAGTGGGATCGGCGGCTTCATCAAGGTGG	545
Db	519	CGACCCGCTCAAGGTGCAACAGCGGCGGCTCAACGGGAGTGGATGTCGTCCTCAACAGCA	578
OY	546	CCAGTACGACCAAGATCCGATCGAGATCTGCGGCAAGAGGCAATCGGCAACGAGTGAAT	605
Db	579	CGCGGCGCGAGGCGCTTCTTCGGGATGCACTGTCAAGCAGGGCGAATTCAGGCGCTGCT	638
OY	606	CGGCGCCACCCCGGTGAACATCATCGGCGCGCAATGTGACCCAGCTGGGCTGACCT	665
Db	639	CATGCGCACCGGCGTGACACCTTCTTCGGCAAGCGCGGACCTGGATGACAGCACAA	698
OY	666	GAATTTCCCATCAGGCCCATCGAGACCGTGGCCCTGTAACTGAAGCCCGGCAATGACGG	725
Db	699	CAACATCGGCACTTCGAGCTCGTGCTCAACGGCATGCGCACTTTCGATCATCTTCAT	758
OY	726	CCCCAAGGTGAAGCAGTGGCCCGTGAACGAGGAAGATCAAGGCGCTGACCGCATCTG	785
Db	759	CGGGGTGCGCATGATTAATCGAGATCATGTTATATATACCGAATCCAGCACCGCGCGTACG	818
OY	786	CGAGAGATGAGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGAGAACCCCTTCAACAC	845
Db	819	CGAGGGCATGAGCAACTGCTGTGCTGTCTCATGCGGGCATGCCCATGCGCATGGCCAC	878
OY	846	CCCCGTGTTGCGCATCAAGAGAGAGACAGCACCAATGGCGGCAAGCTGTGTGAATTCCG	905
Db	879	CGTCTGTCCGTCACTACATGCGCATCGGCTCCCAACCGGCTGTGCGAGAGGGCGCATCAC	938
OY	906	CGAGCTGAACAAGGGGACCCAGGACTTCTGGGAAGTGAAGTGGGCAATCCCGCACCCGCG	965
Db	939	CAAGCGATGACCGGCCATCGAGAGATGCGCGGATGGAAGTGTCTTTCGACGACCAAGAC	998
OY	966	CGGCTTGAAGAGAGAGAGACCGTGAACCGTGTGGAAGTGTGGACGCTTAATTCAAGCT	1025
Db	999	CGGAGCGTGACGCTCAACAAGGCTCAACCGTGAACAAGCGTGATCGAGGTTGTACGGGCG	1058
OY	1026	GCCCTTGAAGAGGACTTCCGCAATGACCGCTTCAACATCCCGACATCAACAACGA	1085
Db	1059	GGGGCTGGAACAAGGACTCGTGTCTCTGTACGCGCGAGAGGCGTCCGCGTGAAGAACCA	1118
OY	1086	GACCCCGCGGACTCGGCTAACAGTACAAAGTGTGCGCCCAAGGCGTGAAGGCGAGCCCGAG	1145
Db	1119	GGAGCGCAATTGACACATGATCATGTGGGAGATGCTGCGGACCCCAAGAGGCGCGCGCGG	1178
OY	1146	CATCTTCCAGAGCAGATGACCAAGATCTTGAGACCTTTCGCGCGCGCAACCCCGAGAT	1205
Db	1179	CATCAAGAGGTCCACTTCTCTCCCGTTC---AACCCTGTGAAGAAAGCGGCAATCAC	1235
OY	1206	CGTATCTTACAGGCGCCCTCTGTACGTGGGCGAGCACTTGAGATCGGCGACGACCGGCG	1265
Db	1236	CTAATCATGACGGCATGTGGCGAATGGCACAGGATCAGAAAGGCGCGCGCGAGCATCAT	1295
OY	1266	CAAGATCGAGAGAGCTTGGGCAAGGACCTGTGCGCTGGGGCTTTCACACCCCGCAAGAA	1325
Db	1296	CGAGCTGTGCAAGATGAGAGAGAGCGCGGAAGAGAGGTGTACACGCTGATCGACAGTA	1355

QY 1011 CGCCTACTTCAAGCGTCCCTTGAAGACAGACTTTCGCAAGTACACCGCTTCACCATCC 1070
 DB 968 CGCCTACATCGCCAGGACAGCGCCCAAGAGCATCGCCAGCGCGCTCGCCCGCGCTG 1027
 QY 1071 CAGCATCAACAGAGACCCCGCGCATCGCTACCAAGTACAACTGTCGCCAGGCTG 1130
 DB 1028 CATCTGACGGGTGTGTGATGACCATCGGCTCCCGACGCGCTTCCTGTTGTGACATC 1087
 QY 1131 GAAGGACAGCCCGACATCTTCCAGAGCAGTACCAAGATCTGAGACCTTCGCGGC 1190
 DB 1088 CTACGGACACCGGACAGATCCCGCAAGAGATCTCTCAAGATCTTCAGAGAACTTGA 1147
 QY 1191 CCGCAACCCCGAGATCTGATCTACAGGCCCCCTGTACGTGGCAGAGCATGAGAT 1250
 DB 1148 TTTCAGGCGCGGAGATGATGACATCACTGACCTCAAGAGGCGCGCAACGGTTTAT 1207
 QY 1251 C 1251
 DB 1208 C 1208

RESULT 4
 LOCUS AY103647 2598 bp mRNA linear HTC 18-FEB-2005
 DEFINITION Zea mays PC042084 mRNA sequence.
 ACCESSION AY103647
 VERSION AY103647.1 GI:21206725
 KEYWORDS HTC.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 2598)
 Gardiner, J., Schroeder, S., Polacco, M. L., Sanchez-Villada, H.,
 Fang, Z., Morgante, M., Landewe, T., Fongler, K., Useche, F.,
 Hanafey, M., Tingey, S., Chou, H., Wang, R., Soderlund, C. and Coe, B. H.
 Jr.
 Anchoring 9,371 maize expressed sequence tagged unigenes to the
 bacterial artificial chromosome contig map by two-dimensional
 overgo hybridization
 Plant Physiol. 134 (4), 1317-1326 (2004)
 2 (bases 1 to 2598)
 Hanley, C. F., Dolan, M., Miao, G. H., Vogel, J. M., Whiteitt, M. S.,
 Arthur, L. W., Hanafey, M., Morgante, M. and Tingey, S. V.
 Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 Unpublished (2002)
 3 (bases 1 to 2598)
 Coe, B. H.
 Direct Submission
 Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 If you are interested in getting corresponding physical clones,
 these are publicly available from ZmDB and may be found by BLAST
 searching at MSU, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Malpoc, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
 Location/Qualifiers
 1..2598
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="MaizeGDB:638378"
 /db_xref="taxon:4577"
 /clone_1lb="Maize Mapping Project/DuPont Consensus
 library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the

overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

Query Match 3.5%; Score 86.8; DB 4; Length 2598;
 Best Local Similarity 42.1%; Pred. No. 2.5e-05;
 Matches 813; Conservative 0; Mismatches 1102; Indels 15; Gaps 5;

QY 296 CCACGACCGGAGCTGACAGGTCGCGGACCAACCCCGGACGAGGCGCGGACG 355
 DB 51 CCACCACTTCCCACTCTTGGCACCAGCGGCCCAACCAACACGCGAGCGACAA 110
 QY 356 GCACAGGACCTTGAATCTTCCCAAGATCACCTGTGGCAAGCGCCCTGTGAGATCA 415
 DB 111 TGGCGGCAACGAGCGCATCTGAGAGACGACCCGCTGAATGAGGCGCGGCGG 170
 QY 416 AGTGGGCGGCAATCAAGAGCCCTGTGACACCGGCGGCAACACCGTGTGG 475
 DB 171 AGTGGCGGAGCCCACTGACGAGTGAAGCGCATGTGGCGGACCGCGGACCGG 230
 QY 476 AGGAGATGAGCTGCGCGGCAAGTGAAGCCCAAGATGATGGCGGCAATCGGCTTCA 535
 DB 231 TGTCAAGATGAGGCTTTCACCTCCGCGTGGCGAAGTGGCCGCTTCGCGCA 290
 QY 536 TCAAGTGGCCAGTACGACGAGATCTGATCGAGATCTGCGGCAAGAGCATCGCA 595
 DB 291 AGGAGCGCTCGGCGTGGCGGTGAGCTGAGAGGAAGCCCGCGCGGTCAAGGCA 350
 QY 596 CGGTGATGAGCGCCCAACCCCGTGAACATCATGTGGCGGCAATGCTTAACCACTGG 655
 DB 351 GCAGGAGATGATCTCTCACTGACATCGCCACGCGGCGGACATCAACGCGCTCACACG 410
 QY 656 GCTGACCTGAATCTTCCCAATCAGCGCCCATCGAGACCGTGGCGTGAAGTGAAGCCG 715
 DB 411 GCTTGGCGGACCTTCCACCGCGCACAGAGAGCGGCGCGCTCCAGTGAAGTTC 470
 QY 716 GCATGACGAGCCCAAGGTGAAGCAGTGGCCCTGACCGGAGAGAAAGATCAAGGCTTGA 775
 DB 471 TCAAGCATCTCAAGCGCGGAATCTTCCGACCGGAGAGAGGAGGACACGCTGCGTGG 530
 QY 776 CGGCATCTGAGAGATGAGAAAGAGGAGAGATCAACCAATCGGCCCGAGAAC 835
 DB 531 AGGTACCCGCGCGGAGTGTGTGCGCATCAACCTCTCCAGGCGTACTCGGCA 590
 QY 836 CCTACAAACCCCGTGTTCGATCAAGAAAGAAAGAGACGACCAAGTGGCGCAAGCTGG 895
 DB 591 TCCGTTGAAGATCTCGAGGCGATCAAGAGCTGTCAACCGGTTCAGGCCCTTGGC 650
 QY 896 TGAATCTCGGAGCTGAACAGCGCACCAAGACTTCTGGAGGTGACGCTGGACATCC 955
 DB 651 TCCGCTCGGAGGACCATCAACCGGCTGGGAGACCTGGTCCGCTTCAATCGGCG 710
 QY 956 CCCACCCCGCGCTTGAAGAAAGAAAGAGGTGACCTGTGTCGACCTGGGCGACGCTT 1015
 DB 711 GCTTCATCAAGGCGCGCCCAACGCGAGCGCGCTGACCG--TCACGAGAGAGGTGG 767
 QY 1016 ACTTCAAGGTGCGCTTGAACAGAGACTTCCGAATACCGGCTTTCACATCCCAAGA 1075
 DB 768 ACGCCGCGAGGCGTTCAAGATGCGCGGCAATGAGAGGCGGCTTTTCAAGCTCAACCCA 827
 QY 1076 TCAACAAAGAACCCCGGACATCCGCTACAGTACCAAGTGTGCGCCCAAGGCTGAAAG 1135
 DB 828 AGGAGGCGCTCGCATCTGCAACGCAAGTCCGTGGGCTCGCGCGTCAACCGTGA 887
 QY 1136 GCAAGCCCAAGATCTTCAAGAGCAGATGACCAAGATCTTGAAGCTTTCGCGCCGCA 1195
 DB 888 TGTACAGGCAACGCTCGCGCGCTGTGTGAGAGTCT--GTCCCGCTTCTGCG 944
 QY 1196 ACCCGAGATGTATCTACAGAGCGCCCTGTAGTGGGAGGACGACCTGAGATCGGC 1255
 DB 945 AGGTATAGAGGCAAGCGCGGATGACGAGCAACCACTGACCAAGCTGAAGCAACC 1004
 QY 1256 AGCAACGCGCAAGATGAGAGCTGCGCAAGACCTGTGCGCTGAGGCTTCAACCAACC 1315

QY 541 GTGCGCAGTACGACCAAGATCTGTATCGA---GATCTGGGCAAGAGCCATCGGACCC 597
DB 641 GAGGTGCGCTACGCGGTGGGCAAGATCTCGAGATCGGCGCCAACTTCTCAACCCGAC 750
QY 598 GTGCTGATCGGCCCCCGGTGAACATCATCGGCGGCAACATCTGAACCCAGCTGGAC 657
DB 751 GTGGATTCGCGGCTCGCTTCCAGAAAGCATCTCAACTCTG----- 798
QY 658 TGGACCCCTGAATCTCCCATCAGCCCCCATTCGAGACCGTCCGTGAAGCTGAAGCCCGAC 717
DB 799 TACATCTCGAGTGAATGAGCTCTCCCGAGGTGCGCACTACTGAGAGAGGATTATCAAG 858
QY 718 ATGACGCGCCCAAGATGAGAGGTGCGCCCTGACCGAGAGAGATCAAGGCGCTGAC 777
DB 859 ATCAACGATACCAAGAGAGCGGTTCTGGAACCGGTGTGTCTGCTCAATGTTCAACAC 918
QY 778 GCCATCTGGAGAGATGAGAGAGGCAAGATCAACAGATCGGCGCCGAGAACCC 837
DB 919 GTGGCGGCAAGAGATCGCGCTGGGCTTCGCTTCAAGAGAGACCGGCGACAC 978
QY 838 TACAACACCCCGGTGTTCGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
DB 979 AGGAG 1038
QY 898 GACTTCGAG 957
DB 1039 ATCTAG 1098
QY 958 CACCCCGCGGCTTGAAG 1017
DB 1099 GATTCGAG 1158
QY 1018 TTCAAG 1077
DB 1159 GTGCTGAG 1218
QY 1078 AACACAG 1104
DB 1219 TGGAG 1245

RESULT 6
LOCUS C2247380 743 bp DNA linear GSS 11-FEB-2005
DEFINITION A1A-asf35a12 b1 Ancylostoma caninum whole genome shotgun library
(A1AAS8 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION C2247380.1 GI:59632821
VERSION GSS.
KEYWORDS Ancylostoma caninum (dog hookworm)
SOURCE Ancylostoma caninum
ORGANISM Ancylostoma caninum (dog hookworm)
REFERENCE 1 (bases 1 to 743)
AUTHORS Milteve,M., McCarter,J.P., Pape,D., Ritter,B., Tsagarelis,H.I., R.,
Runko,I., Martin,J., Wylie,T., Dantle,M., Meyer,R., Messina,D.,
Waterston,R.H., Clifton,S.W. and Wilson,R.
TITLE Genome Survey Sequences from the Parasitic Nematode Ancylostoma
caninum
JOURNAL Unpublished (2004)
COMMENT Contact: Mireya M
Washington University in St. Louis
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: nematode@wustl.wustl.edu
Genomic DNA provided by John Hawdon (mjm@hmgwmc.edu) DNA
sequenced by Washington University Genome Sequencing Center
Classes: shotgun.
FEATURES
source 1..743
/organism="Ancylostoma caninum"

/mol_type="genomic DNA"
/strain="Baltimore"
/db_xref="taxon:29170"
/dev_stage="Adult"
/lab_host="GSI0"
/clone_id="Ancylostoma caninum whole genome shotgun
library (A1AAS8 001)"
/note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
Ancylostoma caninum genomic DNA was randomly sheared,
end-repaired and size fractionated to enrich for 2-4 kb
fragments. Genomic DNA was provided by John Hawdon
(mjm@hmgwmc.edu) at George Washington University.
Sequencing by Washington University Genome Sequencing
Center, St. Louis, MO."

ORIGIN

Query Match 3.5%; Score 86; DB 10; Length 743;
Best Local Similarity 45.9%; Pred. No. 3.2e-05;
Matches 293; Conservative 0; Mismatches 345; Indels 0; Gaps 0;
QY 805 GGCAAGATCAACCAAGATGCGCCCGAGAACCCCTTCAACACCCCGGTTCGATCGCATCAAG 864
DB 6 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 65
QY 865 AAGAAGACAGACCAAGTGGCGCAAGCTGTGATCTTCGCGAGCTGAACAAGCGCAC 924
DB 66 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 125
QY 925 CAGGACTTCTGGAGAGTSCAGCTGGGCAATCCCCACCCCGCGCTGAAGAGAGAGAG 984
DB 126 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 185
QY 985 AGCTGACCGGTGTGAGAGTGGGCGACGCTTACTTACGCTGCCCTTGAAGAGAGATTC 1044
DB 186 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 245
QY 1045 CGCAAGTACACCGCTTCAACCATCCCGAGCATCAACAACAACAACAACAACAACAACA 1104
DB 246 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 305
QY 1105 CAGTACAAAGTGTGCGCCCAAGGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1164
DB 306 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 365
QY 1165 ACCAAGATCTGAGAGCTTCCGCGCCGCAACCCCGAGATGTGATCTTCAACAGGCC 1224
DB 366 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 425
QY 1225 CTGTAGTGGGAG 1284
DB 426 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 485
QY 1285 AAGACCTGTGCGGTGGGCTTCAACACCCCGCAAGAGAGAGAGAGAGAGAGAGAG 1344
DB 486 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 545
QY 1345 TTCTGTGAG 1404
DB 546 GACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 605
QY 1405 CCGAG 643
DB 606 AACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA

RESULT 7
LOCUS BM321451 951 bp mRNA linear BST 03-JAN-2002
DEFINITION rockefeller.0.1211 Mestigameoba balaenuthi lambda ZAP II library
Mestigameoba balaenuthi cDNA similar to adenosylhomocysteinase
3.3.1.1), mRNA sequence.
ACCESSION BM321451
VERSION BM321451.1 GI:18055857


```

Db 453 CGACGTGCG-----CGTGTTCAGCCCGAGAGAGGTGAGCGCCATGTGCTCAC 500
Qy 936 GGAAGGTGACGTGGGCATCCCCACCCCGCCCTGAAAGAAAGAGGTGACCGT 995
Db 501 GCGGATGAAAGAGACGCGCCGAGGCTTACTCGGCAAGAGGTGACCCGCGCTGTGTAC 560
Qy 996 GCTGACGTGGGCGACGCGCTTACTTCAAGCGTCCCTGACGAGACTTCCGCAATGAC 1055
Db 561 CGTCCCGGCTACTTCAAGACGCGCAGCGGCAAGGACCAAGAGACGCGCGTCAATGCG 620
Qy 1056 GCGCTTACCATCCCGAGATCAACAAGAACCCCGGAGATCCGCTACAGTACAGT 1115
Db 621 GCGGCTACCGGTGACCGCATCATCAAGAGCCACCGCGCGCATGTGCTTACGGCAT 680
Qy 1116 GCTGCGCAGGCTGGAAGGAGCGCCAGCATCTTTCAGAGACATGACCAATCTCT 1175
Db 661 CGACAAAGAAAGGCGCGAGAAAGAGTCTGCTTTCAGCTTGGCGGCGACATTTGCA 740
Qy 1176 GAGAGCCTTCCGCGCCCGCAACCCCGAGATCGTGAATCTACAGAGCCCGCTGTAGTG-- 1233
Db 741 CGTGAATCTCTCGCATGACAAAGCGCGTGTTCAGAGTCTTGCACCAAGCGGACAC 800
Qy 1234 ----GCGAGCGACCTGGAGATGCGCCAGACCGCGCCAAAGATCGAGAGCTGCGAAGCA 1289
Db 801 CCACCTCGGCGGCGAGGACTTTCAGCAACGCTCATGACCACTTGTCAAGTATCGG 860
Qy 1290 CCGTGTGCGTGGGCGCTTCAACCCCGCAAGAAAGCAACGAGAGGCGCCCTTCT 1349
Db 861 CGGAAAGCACGCGCGGACATCACCGGCGACCGCGCGCTGGGCGAAGCTCGCGCGCA 920
Qy 1350 GTGATGAGGCTACGAGCTGACCCCGACAAAGTGAACCGTGCAGCCCATGAGCT---GCC 1406
Db 921 GTGCGAGGCGCGCAAGCGCGGCTGACGACCAAGCACAGGTCGCGGTGAGGTGAGTGC 980
Qy 1407 CGAAGAAAGAGAGCTGGAACCTGGAAGCATTCAGAAAGCTGTGGCGAAGCTGAGTGGCC 1466
Db 981 CCGTGTTCAGCGCGCTGCACTTGTGCGAGCGCGCTTCCCGGCGCGCTTTCAGAGACTCA 1040
Qy 1467 CAGCGAGATCTACCCCGGATCAAGGTGCGCAAGTGTGCGAAGCTGTGCGGCGGCGCA 1526
Db 1041 CAGCGACCTCTTTCAGAAAGACATGATGTGCGGTGAGAAAGCCATGTGCGAGAGCGCGCT 1100
Qy 1527 GCGCGCTGACCGACATCTGTGCGCCCTGACGAGAGGCGCGAGCTGAGAGCTGGCGAAGCCG 1586
Db 1101 GAGCAAGGCGGACATGACGAGATGCTCTCGTCCGCGGACAGACCAAGATCCCAAGGT 1160
Qy 1587 CGAGATCTGTGCGGAGCGCGGTGCAAGGCTGTACTACGACCCGAGCAAGACCTGTGTCG 1646
Db 1161 GCGAGAGCTGTCAAGGACTACTTGGGTGCAAGAGCCCAACGCGGCGTCAACCCGCA 1220
Qy 1647 CGAGATCTCAAGAGCGGCGCAAGACAGTGAAGCTTACAGATCTTACAGAGACCTTCA 1706
Db 1221 CGAGGCGGTGCGGTGAGCGCGCGGTGAGGCGCGAGCATTCAGCGGCAAGTGTGACGA 1280
Qy 1707 GAACTGAAAGACCGGCAAGTACGCAAGATGCGCACCGGCGCACCAACAGAGTGAAGCA 1766
Db 1281 GAAACAGGAGAGATGATCTCTGAGCGTCCGCGGCTACCTTGGCGCTTGAAGAGCGCG 1340
Qy 1767 GCTGACCGAGGCGGTGAGAGATGCGATGAGAGACATGTGATCTGGGCGAAGACCCG 1826
Db 1341 CGGCGGCGGTGATGCGCAAGCTGATCCGCGCAACGAGTGTGCGCAAGAAAGCGCA 1400
Qy 1827 CAAGTTCCGCGCTGCCATCAAGAGAGACTGGAGAGCTGTGTGACCGACTTACGTGCA 1886
Db 1401 GGTGTTCACACGATCAAGAGCAAGAGCAACGAGTCAATCAAGTGTTCAGAGGCGCA 1460
Qy 1887 GGCACCTTGATCCCGAGTGGAGTTCGTGAACACCCCGCGCTGTGAGACTGTGTGTA 1946
Db 1461 GCGGAGCATGAGAGAGGACAAACGCGGTGCTCGGCAAGTTCGACTTCGCGGATGCGCGC 1520
Qy 1947 CGAGCTGAGAAAGAGCCCATCATCGCGCGCAAGACTTTCATGTGAGCGGCGCGCA 2006
Db 1521 GCGCGCGAGAGGCGCGCGAGATGAGGTGACG-----TTGAGGTGAGAGCGCA 1571

```

```

Qy 2007 CCGGAGAGCAAGATCGGCAAGCGCGCTAGGTGACGACCGGCGCGCGAGAGATGCT 2066
Db 1572 CGGCATCTCTAGCGGTGTGCGCGCGAGCAAGGCCACCGGCAAGGTGAGAAATCACT 1631
Qy 2067 GAGCTGACCGAGACCAACCAACAGAGACCGAGCTGAGGCGCATCGAGCTGGCGTGA 2126
Db 1632 CTCGCGGCGACGACCGCAAGATCAGCAGAGAGATGACCGGAGTGTGTGCGGAGCGGA 1691
Qy 2127 GAGACGCGGCGAGCGGTGAACATGTGACCGACAGCGATCGCTTGGGATATCA 2186
Db 1692 GAGATTCCCGAGAGAGACCGCGGACAGAGAGAGGTGACGCGCCGAAACAGCTGGA 1751
Qy 2187 GCGCCAGCGCGACAGAGAGAGAGCGAGCTGTGAACAGAT-----CATGACAGCT 2240
Db 1752 GCGGTACGTTCACATCAAGAAACACGCTGCGCGCAAGATGCGGACGCGCATGAGAG 1811
Qy 2241 GATCAAGAGAGAGAGGTGTACTTGAAGCTGTGCGCGCGCCCAAGAGGCAATCGCGCA 2300
Db 1812 CGAGGAGAGAGCAAGGTGAGAGAGCGGTGAGAGGAGCGGTACAGTGGCTGAGCGCA 1871
Qy 2301 CGAGCATTCGACAAAGCTGTGAGCAAGGCGCATCCGCAAGTGTCTTGTGACGCGAT 2360
Db 1872 CCGGACCGCGGCAAGAGAGTACGAGAGAGCTGAGAGGAGCTGTGAGAGCTGTGCA 1931
Qy 2361 CGATGCGGCGCATCGTATCTACCAAGTACAGAGCAACCTGTACGTGAGGCG 2409
Db 1932 CCGGTATGTGCGCGGTCTTACAGAGGTCCGCGCGCGCGCGCGCG 1980

RESULT 9
CL978463 1725 bp DNA linear GSS 21-SEP-2004
LOCUS CL978463
DEFINITION OsIFCC031917 Oryza sativa Expressed Library Oryza sativa (indica
VERSION CL978463.1 GI:52411427
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Bukariyola; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1725)
AUTHORS Ma,L., Wang,J., Chen,C., Liu,X., Su,N., Li,L., Wang,X., Cao,M.,
Jiao,Y., Sun,N., Zhang,X., Bao,J., Sun,D., Zhao,H., Yuan,L.,
Wong,G.K.S., Deng,X.W. and Wang,J.
TITLE An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.
FEATURES
source
1..1725
Location/Qualifiers
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:3946"
/clone_lib="Oryza sativa Expressed Library"
/note="Oryza sativa exon trapped genomic sequences "
ORIGIN
Query Match 3.4%; Score 84; DB 10; Length 1725;
Best Local Similarity 43.7%; Pred. No. 7.8e-05;
Matches 470; Conservative 0; Mismatches 600; Indels 6; Gaps 2;
Qy 949 GGCATCCCGCACCGCGCGCTGAAAGAAAGAGAGGTGACCGTGTGAGCGTGGCG 1008

```


QY 1237 AGGACCTGAGATCGGCGCAAGACCGCGCGCAAGATCGAGAGCTGGCGCAAGACCTGCTG 1296
DB 1822 GTTATGCTTGAAGCGCGCGCAAGCTCCACCTCTCTGCGACATCCCGAGCGCG----- 1876
QY 1297 CGCTGGGGTTTACCAACCCCGCGCAAGAAAGACACAGAAAGAGCGCCCTTCCGTGGAGT 1356
DB 1877 -CGCGCGCGAGCG 1935
QY 1357 GCGTACGAGCTGACCGCGCGCAAGTGAAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 1416
DB 1936 CG 1995
QY 1417 AGCTGACCGTGAAGCAATCCAGAGTGTGTGGGCAAGCTGAATCTGGCGCGCGCGCGCG 1476
DB 1996 GTGCGCGCTCG 2055
QY 1477 TACCGCGCGATCAAGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1536
DB 2056 TACCTGGGCTCTCGCGCAACACCAATGAGCGACCTCCGAGCTCGTCAAGAGCTCCGCG 2115
QY 1537 GACATGTCGCTGACCG 1596
DB 2116 AACCTGCAAGCGTGAAGTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 2175
QY 1597 CGCGAGCGCGTGAAGCGC 1614
DB 2176 CGCAACCTCGCTCGCGC 2193

RESULT 12
CK159167/c 869 bp mRNA linear EST 05-DEC-2003
LOCUS FGAS040564 Triticum aestivum FGAS: TaITS5 Triticum aestivum cDNA,
DEFINITION
CK159167
CK159167.1 GI:38985053
RNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Triticum aestivum (bread wheat)
Triticum aestivum
Bakeryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.

REFERENCE
AUTHORS
1 (bases 1 to 869)
Allard, P., Crosby, W.L., Danyluk, J., Eudes, F., Frick, M., Gaudet, D.,
Genswein, B., Graf, R., Gulick, P., Hyman, L.D., Larocque, A.,
Liu, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
Peniket, C., Roach, J.L. and Sarhan, P.
Functional Genomics of Abiotic Stress in Wheat and Canola Crops
Unpublished (2003)
Contact: Wm L Crosby
Bioinformatics
University of Saskatchewan, Department of Computer Science
1C101 Engineering Building, 57 Campus Drive, Saskatoon,
Saskatchewan, S7N 5A9, Canada
Tel: 306 966 1769
Fax: 306 966 2033
Email: fgas.estes@usask.ca

TITLE
JOURNAL
COMMENT
This sequence is the direct result of the Base calling software
Phred (default parameters). It is the raw base calls. To aid in the
identification of the high quality insert the software Lucy
(default parameters) has been run on this sequence. Lucy identified
the region [128,636].
Plate: TaITS537 row: N column: 23.
Location/Qualifiers

FEATURES
source
1..869
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Wheat line PI 178383"
/db_xref="taxon:4565"
/lab_host="DHS alpha"
/clone_id="Triticum aestivum FGAS: TaITS5"
/note="Organ: Crown; Vector: pGEM-T; SSH (suppression

subtractive hybridization) cDNA library from genotype
PI178383 cold hardened at 2 C for 21 days and 49 days
(equal amount of cDNA pooled together before subtraction,
tester) and subtracted against genotype Norstar cold
hardened at 2 C for 1 day (24 H (driver). Modified Smart
cDNA (Clontech) priming and non-directional cloning"

Query Match 3.3%; Score 80.6; DB 7; Length 869;
Best Local Similarity 44.9%; Pred. No. 0.0003;
Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 580 AAGAGGCGATGCGGACCGTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
DB 827 ATGGCG 768
QY 640 ATGTGACCGACCTGGGCTGACCGCTGACCTTCCCATGACCGCGCGCGCGCGCGCGCGCG 699
DB 767 CACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 708
QY 700 GTGAAGCTGAAGCG 759
DB 707 ACCAAC 648
QY 760 AAGATCAAGCG 819
DB 647 AAC 588
QY 820 ATGGCG 879
DB 587 AAC 528
QY 880 AAGTGGCGGACGCTGTGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 939
DB 527 AAC 468
QY 940 GTGCGAGTGGGATCCCCCG 999
DB 467 AAC 408
QY 1000 GACGTGGGCG 1059
DB 407 GAC 348
QY 1060 TTACCATCCCGACGATCAACACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1119
DB 347 GAC 288
QY 1120 CCCGAGGCTGGAAGGCG 1179
DB 287 AAC 228
QY 1180 CCCTTCG 1239
DB 227 AAC 168
QY 1240 GACCTGAGATCGCGCGCGC 1258
DB 167 AACGCGACACACACACACAC 149

RESULT 13
CZ216254 892 bp DNA linear GSS 10-FEB-2005
LOCUS A11A-aaf23g02.b1 Ancylostoma caninum whole genome shotgun library
DEFINITION (A11AGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
ACCESSION CZ216254
VERSION CZ216254.1 GI:59229909
KEYWORDS GSS.
SOURCE Ancylostoma caninum (dog hookworm)
ORGANISM Ancylostoma caninum
Bakeryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
Ancylostomatoidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.

QY 507 CAAGATGATCGGCGGATCGGCGCTTCAATCAAGGTCCGCAAGTACGACCAAGTCTGAT 566
 DB 405 GGCCTACCTCGGCTCGTCTCAAGAAACCCCGTGTACCGTCCCGCTTACTTCAACGA 464
 QY 567 CGAAGTCTCGGCAAGAGGCAATCGGCAACCGTGTATCGGCGCCACCCCGTGAACAT 626
 DB 465 CTGCGAAGAGGCAAGGCAACCAAGAGCCCGGCTCATCGCGGCTCAAGTATGTGCGAT 524
 QY 627 CATCGGCGCAACATGCTGACCCAGCTGGGCTGACCTTGAATTCCCATGAGCCCAT 686
 DB 525 CATCAACGAGCCACCGCGCGGCTCATCGGTACCGGCTTCAAGAAAGGCGACGAGCA 584
 QY 687 CGAAGCCGTCCCGTGAAGTGAAGCCCGCATGACGCGCCCAAGGTGAAGATGGCC 746
 DB 585 CGCGGAGAAAGTGTCTCATTTTGAACCTCGGCGGCGGCACTTGAATGTGTGCTGCT 644
 QY 747 CCTGACCGAGGAAGATCAAGGCGCTTGAACCGCATCTGCGAAGATGGAAGAGGAG 806
 DB 645 CACCATCGAGAGAGGCACTTTCAGAGTCAAGGCAACCGCGCGGACACCACTTCGAGCG 704
 QY 807 CAAGATCACCAAGATGCGGCGCCGAGAACCTTAC--AACACCCCGTGTTCGCATCA 863
 DB 705 CGAAGCTTTCGACACCGCATGTGTGAACACTTCTGTGAGAGTTCAAGCGCAAGAA 764
 QY 864 GAAAGAGAACAGCAACAAGTGGCGCAAGCTGTGTGACTTCGCGAGCTGAACAAGCGCAC 923
 DB 765 GAAAGCATCAAGGCGCAACCGCGCGGCTGCGGCGCTCAAGGACGCGTGCAGCGCC 824
 QY 924 CAGAGCTTCTGGAGGTGCAAGTGGGCAATCCCGCAACCCCGCGCTGAAGAAAGAA 983
 DB 825 CAAGCGCAAGCTGTCTGACGCGGCAACACCATTTAGANTGACTCGTGTACAGAGG 884
 QY 984 GAGCTGACCGGTCTGAGCGTGGGCGACGCTTACTTCAAGGTGCCCTGTGACGAGACTT 1043
 DB 885 CATGCACTTTCACACGACATACACAGGCGCGGTTTGAGAGCTCAACATGACCTGTT 944
 QY 1044 CCGCAAGTACACCGCTTCCATCCATCCCAAGATCAACAGAGACCCCGCGCATCCGCTTA 1103
 DB 945 CCGCAAGTGCATGAGACCGGTGAGAAAGTCTCCCGACGCGCAAGATGGAACAAGAGCAG 1004
 QY 1104 CCAAGTCAACGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1163
 DB 1005 CGTGACGACGTGTCTGTGTGCGGCGGCTGACACCGGATCCCAAGTGAAGAGCTGCT 1064
 QY 1164 GACCAAGATCTGTGAGCGCTTCCGCGCGGCAACCCGAGATGTATCTAACGAGCGCC 1223
 DB 1065 CCGAGCTTCTTCAAGGCGAGAGCTGTGCA-----AGAGCATCAACCCGACAGGCG 1118
 QY 1224 CCTGTACGTGGGCGAGCACTGTGAGATCGGCGCAGCAACCGCGCAAGATCGAGAGCTGCG 1283
 DB 1119 CGTGGGTACGCGCGCGCGCTCCAGGCGCGCATCTGAGGCGGAGAGGCAAGAGAGT 1178
 QY 1284 CAAGCACTGTCTGCGTGGGCTTTCACACCCCGCAGAAAGACCAAGAGAGAGCGCC 1343
 DB 1179 GCGAGACTGTCTCTCTGACGTCAACCCCGCTGTCTCTGCGCTCGAGACCGCGCGG 1238
 QY 1344 CTTCCTGTGAGTGGGCTGAGAGCTGACCCCGCAAGAGTGAAGCTGTGCA---GCCATGCA 1400
 DB 1239 CGTATGACCGGTCTCTATCCGAGAAACACACATCCGACCAAGAGAGAGAGTCTT 1298
 QY 1401 GCTGCCGAGAGAGAGAGTGAACGTGAAACGATCCAGAACTGTGTGGCAAGCTGAA 1460
 DB 1299 CTCACCTACTCTCGACAAACGCGCGCGCTGTGATCAAGTGTGAAGAGGCGAGCGCG 1358
 QY 1461 CTGGGCGAGCAATCTAACCCCGGCAAGAGTGGCGGCAAGCTGTGCA-----AGCT 1511
 DB 1359 CAGAGACCAAGAGCAACACTCTCTCGGCAAGTTCGAGCTTCGCGCATCCACCGCGCC 1418
 QY 1512 GCTGCGGCGGCGCAAGGCGCTGACCGACGATGTGTGCGCTGACGAGAGGCGGAGCTGGA 1571
 DB 1419 GCGCGGCGTCCCGCATCAAGTGTGTCTTCAATGACGCGCAACGCGCATCTCTGAAGCT 1478

QY 1572 GCTGCCGAGAACCGGAGATCTCGCGGAGCGCCGCGGCAACGCGGTGTACTGACACCCAG 1631
 DB 1479 GTCCGCGAGAGCAAGAGCAACGCGGCGAGAAACAAATACATCAACAGACAAAGG 1538
 QY 1632 CAAGACCTGTGTGCGGAGATCCAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1691
 DB 1539 GCGGCTGAGCAAGAGAGATGAGAGATGTGTGAGAGGCGGAGAGTACAAAGGCGCG 1598
 QY 1692 CAGAGACCTTTCAGAACTTGAAGACCGGCAAGTACGCGCAAGATGCGGCGGCGGCGG 1751
 DB 1599 GAGCGAGAGCAAGAGAAAGTGTGACCGCAAGACGCGCTTCGAGATCTAGCCCTTCA 1658
 QY 1752 CAACGACGTGAGACGCTGACCGAGGCGGTGAGAAAGATGCGCATGAGAGATGTGAT 1811
 DB 1659 CATGCGCAACACATCAAGAGAGAGAGATGCGCTTCAGAGTGTGCGCGGCGGCAAGAA 1718
 QY 1812 CTGGGCGAAGACCCCAAGTTCCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1871
 DB 1719 GAGATTCAGAGAGCGCATGATGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1778
 QY 1872 GACCGACTATCTGAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1931
 DB 1779 GCGCGAGAGTTCAGAGCA---AGATGAAGAGAGTGTGAGGCGCATCTGCAACCGCATCAT 1835
 QY 1932 GGTGAAGCTGTGTGTACCAAGTGTGAGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1991
 DB 1836 CCGCAAGATGTACAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1895
 QY 1992 GAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2036
 DB 1896 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1940

RESULT 15
 BM320864
 LOCUS
 DEFINITION
 Masticamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
 sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Masticamoeba balamuthi
 Masticamoeba balamuthi
 Balamuthi, P., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W.,
 Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M. and
 Philippe, H.
 The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Masticamoeba
 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 11830664
 COMMENT
 Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockefeller.edu
 Insert Length: 1132 Std Error: 0.00
 POLYA=No.

FEATURES
 source
 1. 1132 Location/Qualifiers
 /organism="Masticamoeba balamuthi"
 /mol_type="mRNA"
 /strain="ATCC 30984"
 /db_xref="taxon:108607"
 /clone_lib="Masticamoeba balamuthi lambda ZAP II library"
 /note="syn: Phreatamoeba balamuthi"

ORIGIN
 Query Match 3.3%; Score 80.2; DB 3; Length 1132;
 Best Local Similarity 45.3%; Pred. No. 0.00037;
 Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

QY 139 CGCGCCCCCGAAGAGGCTGTGAAAGTCGCGAAGAGGCGCACCAATGAAGAC 198
DB 79 CGCCCGCAGAGGCGAAGCGGACTACCGCGCGCGCCANCTGTATCCAGAGCAAGAACAG 138
QY 199 TGCACCGAGCGCGCAAGCCCACTTCTTCGCGAGAGACTGCGCTTCCCGCAGAGCGC 258
DB 139 TACCAACAGCGCCCAAGTACCGCTTCGTC-----GTCCGCTTCAACCAAGGAGCATCGTC 192
QY 259 CGCGAGTTCCCGAGGAGAGAACCGCGCCCAAGCCCCCAAGCGCGAGCTGCAGGTC 318
DB 193 TGCAGATGCGCTTACCGCCCAAGATCGACGCGAGCACATCTGCGCGCGCTACTCGCAC 252
QY 319 CGCGGCGAACAACCCCGCAGCGAGCGCGCGCGCGCGCGCAAGCGCACTTCCCGC 378
DB 253 GAGCTCACCGCGCTTCCGCGTCAAGCTCGGCTGACCACTACGCGCGCGCTACCGGACT 312
QY 379 CAGATCACTCTGTGCGAGCGCGCGCTGTGAGCATCAAGTGGCGCGCGCATCAAGAG 438
DB 313 GCGCTGTGCTGCGCGCGCGCTGTGAGAGGCTCAACTCCAACTACGAGGCT 372
QY 439 GCGCTGTGAGACACCGCGCGCGCGCATCGTGTGAGAGAGATGAGCTGCGCGCGCAAG 498
DB 373 GTCAAGAGAGGTCAACGCGCGAGACTTACAGCTGAGAGAGCTGACGACGCGCGCGCGT 432
QY 499 TCGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAGTGGCGCGCATACGACAG 558
DB 433 TTCAAGGCGCTGTCTGACGTGCGCGCTGTGCGGACCTGACGCGCGCGCGGTTCGCGC 492
QY 559 ATCTGATGAGATCTGCGCGCAAGAGCGCATCGGCACTGTGCTGATGCGCGCGCACCGCC 618
DB 493 GCGCTCAAGGCGATGTGCGACGCGCGCGGTCAAGTCCCG---CACAGCGAGACCGGCTTC 549
QY 619 GTGAACATCATGTGCGCGCAATGTGACCCGAGCTGCGCTGCACTTCCCGCATC 678
DB 550 GTGCGCTTCAACGCGCGCAAGAGAGGCTCAACGCGCGGTCTTCGCAAGTACATCTTC 609
QY 679 AGCCCCATGAGACCGTGCCTGTGAAGCTGAAGCCCGCGCATGAGCGCGCCCAAGGTGAAG 738
DB 610 GCGCGCGCATGTGCGCGCGTACATGAAGCTCTCAAGAGAGAGAGCGCGCGCGCTTCGAC 669
QY 739 CAGTGCCTCTGACCGAGAGAGATCAAGCGCTGACCGCATCTGCGAGAGAGATGAG 798
DB 670 CCGCAGTTCTCGCGCTACCGCAAGAGGAGGTGTCAACGCGCATGTCTGAGAGATCTAC 729
QY 799 AAGAGGCGCAATCAAGATCGCGCGCGCGCAAGCCCTACACACCGCGGTTCGCGC 858
DB 730 ACCGAGGCGCCACAGAGATCGCGCGCGACCGGACTTGTCCCAAGCGCGCTCGAAG 789
QY 859 ATCAAGAGAGAGAGAGAGCAAGCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAG 918
DB 790 CCGAGGCGCGCAAGCGCGCAAGCATGTGGGCAAGCGCGCATGACACAGAGCGCAAG 849
QY 919 CGCACCCAGAGATTCTGAGAGGTGACGCTGGGATCTCCCGACCGCGCGC 966
DB 850 AACCGGTGCGCGCAAGAGAGTCCGCTGAGGCTACCGCGCGCGCGC 897

Search completed: December 31, 2005, 02:31:26
Job time : 8703.67 sec

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: December 30, 2005, 07:51:40 ; Search time 11926.2 Seconds
(without alignments)
11710.708 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457
Sequence: 1 gtcgacgcaccacatgcccga.....gggctagcaccgctgaattc 2457

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 588314 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_stc:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_hcg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	AX455916 Sequence
2	2436.2	99.2	2463	6	AX455915 Sequence
3	2415.4	98.3	2469	6	AX455914 Sequence
4	2040	83.0	2300	6	BD263705 Improved
5	2040	83.0	2300	6	CO870575 Sequence
6	2040	83.0	2300	6	AR373388 Sequence
7	2019.2	82.2	2306	6	BD263704 Improved
8	2019.2	82.2	2306	6	CO870574 Sequence
9	2019.2	82.2	2306	6	AR373387 Sequence
10	2000.8	81.4	2166	6	AX427930 Sequence
11	1998.4	81.3	2312	6	BD263706 Improved
12	1998.4	81.3	2312	6	CO870576 Sequence
13	1998.4	81.3	2312	6	AR373389 Sequence
14	1978.4	80.5	9788	6	AX427936 Sequence
15	1978.4	80.5	9169	6	AX427931 Sequence
16	1975	80.3	9194	6	AX427926 Sequence
17	1973.4	80.3	9194	6	AX427925 Sequence
18	1973.4	80.3	12411	6	AX427927 Sequence

19	1970	80.2	9785	6	AX427938 Sequence
20	1966.6	80.0	9167	6	AX427933 Sequence
21	1966.6	80.0	9170	6	AX427928 Sequence
22	1966.6	80.0	9782	6	AX427935 Sequence
23	1966.6	80.0	9783	6	AX427934 Sequence
24	1966.6	80.0	9790	6	CS078943 Sequence
25	1966.6	80.0	9792	6	AX427932 Sequence
26	1966.6	80.0	36066	6	CS078956 Sequence
27	1965	80.0	9189	6	AX427921 Sequence
28	1962	79.9	4353	11	AF287353 Synthetic
29	1945.4	79.2	3015	6	AX455946 Sequence
30	1928.8	78.5	3009	6	AX455987 Sequence
31	1915.8	78.0	4319	6	BD263639 Improved
32	1915.8	78.0	4319	6	CO870498 Sequence
33	1915.8	78.0	4319	6	AR373382 Sequence
34	1905.2	77.5	4352	11	AF287352 Synthetic
35	1905.2	77.5	4352	11	AF789509 Synthetic
36	1878.6	76.5	1965	6	AX455952 Sequence
37	1872.8	76.2	2299	6	BD263703 Improved
38	1872.8	76.2	2299	6	CO870573 Sequence
39	1872.8	76.2	2299	6	AR373386 Sequence
40	1859.8	75.7	2577	6	AX457088 Sequence
41	1852.2	75.4	4343	6	AX306428 Sequence
42	1852	75.4	2305	6	BD263702 Improved
43	1852	75.4	2305	6	CO870572 Sequence
44	1852	75.4	2305	6	AR373385 Sequence
45	1848.2	75.2	4341	6	AX306429 Sequence

ALIGNMENTS

RESULT 1
LOCUS AX455916 2457 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 32 from Patent WO0204493.
ACCESSION AX455916
VERSION AX455916.1 GI:21714901

KEYWORDS
SOURCE
ORGANISM
other sequences; artificial sequences.

REFERENCE
AUTHORS
TITLES
JOURNAL
CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES
source
1..2457
/organism="synthetic construct"
/mol_type="unasigned DNA"
/db_xref="taxon:32630"
/note="PR975YMMW"

ORIGIN

Query Match 100.0%; Score 2457; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 8.9e-203;
Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GTCGACGCCACATGCGCCGACATGACCGCCACGAGCCCAATCTCGATGCGAG 60
|||||
1 GTCGACGCCACATGCGCCGACATGACCGCCACGAGCCCAATCTCGATGCGAG 60
|||||
61 CCGAGCACTTCAAGGGGCCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCG 120
|||||
61 CCGAGCACTTCAAGGGGCCCCCAAGCGCATCATCAAGTCTTCAACTGCGGCAAGAGGCG 120
|||||
121 CACATGCGCCGCAACTGCGCGCGCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG 180
|||||
121 CACATGCGCCGCAACTGCGCGCGCGCCGCAAGAGGGCTGCTGGAAGTGGCGCAAGAG 180
|||||
181 GCGCACCAATGAAGACTGCAACCGAGCGCCAGGCCCAACTTCTTCGCGCAAGACTGGCC 240
|||||

Db 181 GGCACAGATGAAGACTGCAACGAGGCGCAGGCAACTCTTCGCGAGGACCTGCGC 240
Qy 241 TTCCCCCAGAGGCAAGGCCCCCGAGATTCCCCCAGCAGAGAACCGGGCCCAACGCCCCAC 300
Db 241 TTCCCCCAGAGGCAAGGCCCCCGAGATTCCCCCAGCAGAGAACCGGGCCCAACGCCCCAC 300
Qy 301 AGCCGAGAGCTCAGAGTGCAGGCGCAGCAACCCCGCAGAGAGGCGGCGCGAGGCGCAG 360
Db 301 AGCCGAGAGCTCAGAGTGCAGGCGCAGCAACCCCGCAGAGAGGCGGCGCGAGGCGCAG 360
Qy 361 GGCACCTTGAACTTCCCCCAGATCAACCTGTGAGCAGGCGCCCTGTGAGCATCAAGGTG 420
Db 361 GGCACCTTGAACTTCCCCCAGATCAACCTGTGAGCAGGCGCCCTGTGAGCATCAAGGTG 420
Qy 421 GGGGGCCAGATCAAGAGAGGCGCTGTGCAACCGGCGCGAGACACCGGTGTGAGAGAG 480
Db 421 GGGGGCCAGATCAAGAGAGGCGCTGTGCAACCGGCGCGAGACACCGGTGTGAGAGAG 480
Qy 481 ATGAGCTTGCCCGGCAAGTGAAGGCCCAAGATGATGCGCGCATGCGAGGCTTCAACAAG 540
Db 481 ATGAGCTTGCCCGGCAAGTGAAGGCCCAAGATGATGCGCGCATGCGAGGCTTCAACAAG 540
Qy 541 GTGCGCAGATCAGACCAAGATCTGTATCGAAGATCTGCGCAGAGAGCCATTCGCAACCTG 600
Db 541 GTGCGCAGATCAGACCAAGATCTGTATCGAAGATCTGCGCAGAGAGCCATTCGCAACCTG 600
Qy 601 CTGATCTGCGCCCAACCCCGTGAACATCATTCGGCCGCAATCTGTACCCAGCTGCGCTGC 660
Db 601 CTGATCTGCGCCCAACCCCGTGAACATCATTCGGCCGCAATCTGTACCCAGCTGCGCTGC 660
Qy 661 ACCCTGAATCTTCCCCATCAGCCCATCAGACCGTGCAGGCGCGTGAAGCTGAAGCCCGGCA 720
Db 661 ACCCTGAATCTTCCCCATCAGCCCATCAGACCGTGCAGGCGCGTGAAGCTGAAGCCCGGCA 720
Qy 721 GACGCGCCCAAGAGTGAAGAGTGCCTTGCACCGAGAGAGATCAAGGCCCTTGAACCGCC 780
Db 721 GACGCGCCCAAGAGTGAAGAGTGCCTTGCACCGAGAGAGATCAAGGCCCTTGAACCGCC 780
Qy 781 ATCTCGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATTCGGCCCGGAGAACCTTAC 840
Db 781 ATCTCGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATTCGGCCCGGAGAACCTTAC 840
Qy 841 AACACCCCCGTGTTCGACATCAAGAGAGAGCAGACCAAGTGGCGCAGGCTGTGAGC 900
Db 841 AACACCCCCGTGTTCGACATCAAGAGAGAGCAGACCAAGTGGCGCAGGCTGTGAGC 900
Qy 901 TTCCGAGAGCTGAACAAGCGCACCCAGAGACTTCTGAGAGTGCAGCTGCGCATTCGCCAC 960
Db 901 TTCCGAGAGCTGAACAAGCGCACCCAGAGACTTCTGAGAGTGCAGCTGCGCATTCGCCAC 960
Qy 961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGCAGCGTGTGAGAGCGGCGCATCTTC 1020
Db 961 CCGCGCGGCTGAAGAGAGAGAGAGCGTGCAGCGTGTGAGAGCGGCGCATCTTC 1020
Qy 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTCAACCGCTTCAATCCCAAGCATCAAC 1080
Db 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTCAACCGCTTCAATCCCAAGCATCAAC 1080
Qy 1081 AACAGAGACCCCGGCAATTCGCTACAGTCAACGTCGTCGCCAGGGCTGAGAGGGCAGC 1140
Db 1081 AACAGAGACCCCGGCAATTCGCTACAGTCAACGTCGTCGCCAGGGCTGAGAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAACCTTCCGCGCCGCAACCCC 1200
Db 1141 CCCAGCATCTTCCAGAGCAGCATGACCAAGATCTGAGAACCTTCCGCGCCGCAACCCC 1200
Qy 1201 GAGATCTGATCTACAGAGGCGCCCTGTACGTGGGAGAGGAGCCTGAGATTCGGCAGCAGC 1260
Db 1201 GAGATCTGATCTACAGAGGCGCCCTGTACGTGGGAGAGGAGCCTGAGATTCGGCAGCAGC 1260
Qy 1261 CGCGCCAGAGATCAGAGAGCTGCGCAAGCCTGCTGCGCTGGGCTTCAACACCCCGCAG 1320
Db 1261 CGCGCCAGAGATCAGAGAGCTGCGCAAGCCTGCTGCGCTGGGCTTCAACACCCCGCAG 1320

Db 1261 CGCGCCAGAGATCAGAGAGCTGCGCAAGCCTGCTGCGCTGGGCTTCAACACCCCGCAG 1320
Qy 1321 AAGAGAGCATCAAG 1380
Db 1321 AAGAGAGCATCAAG 1380
Qy 1381 GTGCAAGCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1381 GTGCAAGCCATCGAGCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy 1441 GTGGGCAAGCTGAATCTGGGCGCAGAGATCTAACCCCGCATCAAGGTGCGCTGTGC 1500
Db 1441 GTGGGCAAGCTGAATCTGGGCGCAGAGATCTAACCCCGCATCAAGGTGCGCTGTGC 1500
Qy 1501 AAGCTGTGCGCGGCGCAGAGGCGCTTGAACCAATCGTGCCTTCAACGAGAGAGGCGCAG 1560
Db 1501 AAGCTGTGCGCGGCGCAGAGGCGCTTGAACCAATCGTGCCTTCAACGAGAGAGGCGCAG 1560
Qy 1561 CTGAGCTGCGCGAGAGACCGCGAGATCTTGCAGAGCCCGTGCACGCGGTGTACTACAG 1620
Db 1561 CTGAGCTGCGCGAGAGACCGCGAGATCTTGCAGAGCCCGTGCACGCGGTGTACTACAG 1620
Qy 1621 CCGAGCAAGAGACTGTGTGCGAGATCCAGAGAGAGGCGCAGCAGCAAGTGAACCTACAG 1680
Db 1621 CCGAGCAAGAGACTGTGTGCGAGATCCAGAGAGAGGCGCAGCAGCAAGTGAACCTACAG 1680
Qy 1681 ATCTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1740
Db 1681 ATCTACAGAGAGCCCTTCAAGAACCTTGAAGACCGGCAAGTACGCAAGATGCGCACCGCC 1740
Qy 1741 CACACCAAGCAGAGTGAAGCAGTGCACGAGGCGGTGCAGAGAGATCGCATGAGAGCATC 1800
Db 1741 CACACCAAGCAGAGTGAAGCAGTGCACGAGGCGGTGCAGAGAGATCGCATGAGAGCATC 1800
Qy 1801 GTGATCTGCGGCGAGAGACCCCGAAGTTCCGCTGCCTCATCAGAGAGAGAGACTGCGAGAC 1860
Db 1801 GTGATCTGCGGCGAGAGACCCCGAAGTTCCGCTGCCTCATCAGAGAGAGAGACTGCGAGAC 1860
Qy 1861 TGTGTGAACGACTATGCGAGAGCGCAGTGCAGTCCCGAGTGAAGTTGTGTAACACCCC 1920
Db 1861 TGTGTGAACGACTATGCGAGAGCGCAGTGCAGTCCCGAGTGAAGTTGTGTAACACCCC 1920
Qy 1921 CCCCTGTGAAGCTGTGTACCAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 CCCCTGTGAAGCTGTGTACCAAGCTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1981 TACGTGAACGCGCGCGCAACCGCGAGACCAAGATTCGCGAAGCGCGCTACGTGACCGAC 2040
Db 1981 TACGTGAACGCGCGCGCAACCGCGAGACCAAGATTCGCGAAGCGCGCTACGTGACCGAC 2040
Qy 2041 CCGGCGCGCGCAGAGAGATGTGTAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGCGAG 2100
Db 2041 CCGGCGCGCGCAGAGAGATGTGTAGCTTGAACCGAGACCAACCAAGAGACCGAGCTGCGAG 2100
Qy 2101 GGCATTCAGAGCTGCGCTGAGAGAGCAGCGGCGAGAGTGAACATGTGTGACGAGCGCAG 2160
Db 2101 GGCATTCAGAGCTGCGCTGAGAGAGCAGCGGCGAGAGTGAACATGTGTGACGAGCGCAG 2160
Qy 2161 TACGCGCTGTGAGCATCAAGGCCAGCCGACCAAGAGAGAGAGAGAGAGAGTGTGAACAG 2220
Db 2161 TACGCGCTGTGAGCATCAAGGCCAGCCGACCAAGAGAGAGAGAGAGAGAGTGTGAACAG 2220
Qy 2221 ATCATTCAGAGCTGTATCAAGAGAGAGAGAGAGAGTGTACCTGAGCTGTGGTCCCGCCAG 2280
Db 2221 ATCATTCAGAGCTGTATCAAGAGAGAGAGAGAGAGTGTACCTGAGCTGTGGTCCCGCCAG 2280
Qy 2281 GGCATTCGCGGCGCAAGAGAGATGCAAGAGCTGTGAGAGAGAGAGAGAGAGTGTGAG 2340
Db 2281 GGCATTCGCGGCGCAAGAGAGATGCAAGAGCTGTGAGAGAGAGAGAGAGAGTGTGAG 2340
Qy 2341 TTCTGTGAACGCGCATGATGCGCGCATGTGTACTACAGTACATGAGACGACTGTACGTTG 2400
Db 2341 TTCTGTGAACGCGCATGATGCGCGCATGTGTACTACAGTACATGAGACGACTGTACGTTG 2400

QY 2401 GGCAGCGGGGCGCCCTGATGATTCGATTTAAAGCTTCCCGGGGCTTAGCAGCCGGTGAATTC 2457
Db 2401 GGCAGCGGGGCGCCCTGATGATTCGATTTAAAGCTTCCCGGGGCTTAGCAGCCGGTGAATTC 2457

RESULT 2
AX455915
LOCUS AX455915 2463 bp DNA linear PAT 06-JUL-2002
DEFINITION Sequence 31 from Patent WO0204493.
ACCESSION AX455915
VERSION AX455915.1 GI:21714900
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1
AUTHORS zur Megede, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, B.
TITLE Polynucleotides encoding antigenic hiv type c polypeptides,
JOURNAL Patent: WO 0204493-A 31 17-JAN-2002;
FEATURES CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)
source 1. 2463
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="PR975YM"

ORIGIN
Query Match 99.2%; Score 2436.2; DB 6; Length 2463;
Best Local Similarity 99.6%; Pred. No. 5.5e-201;
Matches 2454; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 GTGAGCGCCACCATGCGCCGAGGCGCATGAGCGCAGCCAGCCAGCAATCTGATGAG 60
Db 1 GTGAGCGCCACCATGCGCCGAGGCGCATGAGCGCAGCCAGCCAGCAATCTGATGAG 60
QY 61 CGGAGCACTTCAGAGGGGCCCGCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC 120
Db 61 CGGAGCACTTCAGAGGGGCCCGCAAGCGCATCATCAAGTGTCTTCAATGCGGCAAGAGGGC 120
QY 121 CACATGCGCCGCACTGCGCGCGCGCGCGCGCGAGAGGGCTGTGAAAGTGCAGAGAG 180
Db 121 CACATGCGCGCAACTGCGCGCGCGCGCGCGCGAGAGGGCTGTGAAAGTGCAGAGAG 180
QY 181 GGCACCAAGTGAAGGACTGCAACGAGCGCAGGCCAATCTTCTCGCGAGGACTGGCC 240
Db 181 GGCACCAAGTGAAGGACTGCAACGAGCGCAGGCCAATCTTCTCGCGAGGACTGGCC 240
QY 241 TTCCGCCAGAGGAGCGCGCGAGTTCCCGAGGAGAGAACCGGCGCAACAGCCCGAC 300
Db 241 TTCCGCCAGAGGAGCGCGCGAGTTCCCGAGGAGAGAACCGGCGCAACAGCCCGAC 300
QY 301 AGCGGAGAGTGTGAGGTGCGCGCGAGCAACCCCGCAGAGGGCCCGCGCGCGAG 360
Db 301 AGCGGAGAGTGTGAGGTGCGCGCGAGCAACCCCGCAGAGGGCCCGCGCGCGAG 360
QY 361 GGCACCTGAACCTTCCCGCAGATCACTGTGAGAGGGCCCGCTGTGAGACATCAAGGT 420
Db 361 GGCACCTGAACCTTCCCGCAGATCACTGTGAGAGGGCCCGCTGTGAGACATCAAGGT 420
QY 421 GCGGCGCAAGTCAAGAGAGGCGCTGTGAGCAACCGGCGCAGAGCAACCGTGTGAGAG 480
Db 421 GCGGCGCAAGTCAAGAGAGGCGCTGTGAGCAACCGGCGCAGAGCAACCGTGTGAGAG 480
QY 481 ATGAGCTGTGCGCGCAAGTGAAGCCCAAGATGATGTGGCGCATGTGGCGCTTCATCAAG 540
Db 481 ATGAGCTGTGCGCGCAAGTGAAGCCCAAGATGATGTGGCGCATGTGGCGCTTCATCAAG 540
QY 541 GTGCGCAAGTCAAGAGCAAGATCTGATGAGATCTGCGGCAAGAGGCCATCGGCAACCGT 600
Db 541 GTGCGCAAGTCAAGAGCAAGATCTGATGAGATCTGCGGCAAGAGGCCATCGGCAACCGT 600

QY 601 CTGATCGGCCCAACCCCGCTGAAATCATATGCGCGCGCAAGATGTGACCCGAGCTGGCTGC 660
Db 601 CTGATCGGCCCAACCCCGCTGAAATCATATGCGCGCGCAAGATGTGACCCGAGCTGGCTGC 660
QY 661 ACCCTGAATCTTCCCATCAGCCCGCATCGAGACCGTGCCTGTAAGTGTGAAGCCCGCATG 720
Db 661 ACCCTGAATCTTCCCATCAGCCCGCATCGAGACCGTGCCTGTAAGTGTGAAGCCCGCATG 720
QY 721 GAACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGAGAAATCAAGGCGCTGACCGCC 780
Db 721 GAACGCGCCCAAGGTGAAGAGTGGCCCTGACCGAGAGAAATCAAGGCGCTGACCGCC 780
QY 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAAATGTGGCCCGCGAACCCTTAC 840
Db 781 ATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAAATGTGGCCCGCGAACCCTTAC 840
QY 841 AACACCCCGGTGTCGCGCATCAAGAAAGAGAGCAGACCAAGGTGGCGCAAGCTGTGAGAC 900
Db 841 AACACCCCGGTGTCGCGCATCAAGAAAGAGAGCAGACCAAGGTGGCGCAAGCTGTGAGAC 900
QY 901 TTCCGAGACTGAACAAGCGCACCCAGAGACTTCTGAGAGGTGAGCTGTGGCATCCCGCAC 960
Db 901 TTCCGAGACTGAACAAGCGCACCCAGAGACTTCTGAGAGGTGAGCTGTGGCATCCCGCAC 960
QY 961 CCGCGCGCTGTGAAGAAAGAAAGAGCGTGAACGCTGTGAGCTGTGGCGCGACCTTACTTC 1020
Db 961 CCGCGCGCTGTGAAGAAAGAAAGAGCGTGAACGCTGTGAGCTGTGGCGCGACCTTACTTC 1020
QY 1021 AGGTGCGCTGTGAAGAGACTTCCGCAAGTCAACCGGCTTGCATCCCGCAGATCAAC 1080
Db 1021 AGGTGCGCTGTGAAGAGACTTCCGCAAGTCAACCGGCTTGCATCCCGCAGATCAAC 1080
QY 1081 AACGAGACCCCGGCGCATTCGCTTACAGTACAAGTCTGCGCGCGGCTGGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCGCATTCGCTTACAGTACAAGTCTGCGCGCGGCTGGAAGGGCAGC 1140
QY 1141 CCGAGCATTTTCAAGAGCAGCATGACCAAGATCTTGAAGCCTTCCGCGCGCGCAACCCC 1200
Db 1141 CCGAGCATTTTCAAGAGCAGCATGACCAAGATCTTGAAGCCTTCCGCGCGCGCAACCCC 1200
QY 1201 GAGATGATCTACAGAGCGCCCGCTGTAGTGGAGAGGAGCCTTGAAGTGGCGAGCAC 1260
Db 1201 GAGATGATCTACAGAGCGCCCGCTGTAGTGGAGAGGAGCCTTGAAGTGGCGAGCAC 1260
QY 1261 CGGCGCAAGTGAAGAGTGTGCGCAAGCACTGTGCTGCTGGGCTTCAACACCCCGAC 1320
Db 1261 CGGCGCAAGTGAAGAGTGTGCGCAAGCACTGTGCTGCTGGGCTTCAACACCCCGAC 1320
QY 1321 AAGAGCACCAAGAGAGCGCCCTTCTGCGCAT-----GAGCTGCAACCCGACAG 1374
Db 1321 AAGAGCACCAAGAGAGCGCCCTTCTGCGCAT-----GAGCTGCAACCCGACAG 1374
QY 1375 TGAACCGTCAAGCCCATGAGCTGCGCGAGAGAGAGTGTGACCGTGAACGATCTCAG 1434
Db 1375 TGAACCGTCAAGCCCATGAGCTGCGCGAGAGAGAGTGTGACCGTGAACGATCTCAG 1434
QY 1435 AAGCTGTGGGCAAGTGAAGTGTGAGCAACGAGATCTAACCCCGGCAATCAAGGTGCGCAG 1494
Db 1435 AAGCTGTGGGCAAGTGAAGTGTGAGCAACGAGATCTAACCCCGGCAATCAAGGTGCGCAG 1494
QY 1441 AAGCTGTGGGCAAGTGAAGTGTGAGCAACGAGATCTAACCCCGGCAATCAAGGTGCGCAG 1500
Db 1441 AAGCTGTGGGCAAGTGAAGTGTGAGCAACGAGATCTAACCCCGGCAATCAAGGTGCGCAG 1500
QY 1495 CTGTGCAAGCTGTGCGCGCGCGCAAGCGCTTGAACGATCTGTGCGCTTGAACGAGAG 1554
Db 1495 CTGTGCAAGCTGTGCGCGCGCGCAAGCGCTTGAACGATCTGTGCGCTTGAACGAGAG 1554
QY 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGCGCTTGAACGATCTGTGCGCTTGAACGAGAG 1560
Db 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGCGCTTGAACGATCTGTGCGCTTGAACGAGAG 1560
QY 1555 GCGGAGTGAAGTGTGCGCGCGAGAACCGCGAGATCTGTGCGGAGCGCGTGAACGCGGTGAC 1614
Db 1555 GCGGAGTGAAGTGTGCGCGCGAGAACCGCGAGATCTGTGCGGAGCGCGTGAACGCGGTGAC 1614
QY 1561 GCGGAGTGAAGTGTGCGCGAGAACCGCGAGATCTGTGCGGAGCGCGGTGAC 1620
Db 1561 GCGGAGTGAAGTGTGCGCGAGAACCGCGAGATCTGTGCGGAGCGCGGTGAC 1620
QY 1615 TACGACCCAGCAAGAGCACTGTGTGCGGAGATCTCAAGAGAGGCGCAAGCATGTGAGCC 1674
Db 1615 TACGACCCAGCAAGAGCACTGTGTGCGGAGATCTCAAGAGAGGCGCAAGCATGTGAGCC 1674
QY 1621 TACGACCCAGCAAGAGCACTGTGTGCGGAGATCTCAAGAGAGGCGCAAGCATGTGAGCC 1680
Db 1621 TACGACCCAGCAAGAGCACTGTGTGCGGAGATCTCAAGAGAGGCGCAAGCATGTGAGCC 1680

QY 1675 TACGAGATCTACGAGAGCCCTTCAAGACTGAAAGACCGGACAGTACGCCAAGATGCGC 1734
 DB 1681 TACGAGATCTACGAGAGCCCTTCAAGACTGAAAGACCGGACAGTACGCCAAGATGCGC 1740
 QY 1735 ACCGCCACACCAAGAGCTGAGAGCTGACCGAGGCGCTGACAGAGATGCGCATGAG 1794
 DB 1741 ACCGCCACACCAAGAGCTGAGAGCTGACCGAGGCGCTGACAGAGATGCGCATGAG 1800
 QY 1795 AGCATGTGATCTGGGGGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAGAGACTG 1854
 DB 1801 AGCATGTGATCTGGGGGCAAGACCCCAAGTTCGCGCTGCGCATCCAGAGAGACTG 1860
 QY 1855 GAGACTGTGAGACCGACTGAGAGCGCACTGAGATCCCGAGTGGAGTTCTGTAA 1914
 DB 1861 GAGACTGTGAGACCGACTGAGAGCGCACTGAGATCCCGAGTGGAGTTCTGTAA 1920
 QY 1915 ACCCCCCCTGTGTGAGCTGTGTGACCAAGCTGAGAGAGAGCCCATCTGCGCGCGAG 1974
 DB 1921 ACCCCCCCTGTGTGAGCTGTGTGACCAAGCTGAGAGAGAGCCCATCTGCGCGCGAG 1980
 QY 1975 ACCCTTACGTGAGACCGGCGCGCAACCGGAGACCAAGATGCGGAGAGCGGCTAAGT 2034
 DB 1981 ACCCTTACGTGAGACCGGCGCGCAACCGGAGACCAAGATGCGGAGAGCGGCTAAGT 2040
 QY 2035 ACCGACCGGCGCGGACAGAGATCTGTAGAGCTGACCGAGACCAACCAAGAGACCGAG 2094
 DB 2041 ACCGACCGGCGCGGACAGAGATCTGTAGAGCTGACCGAGACCAACCAAGAGACCGAG 2100
 QY 2095 CTGCAAGGCGATCAAGCTGCGCTGTGACAGAGCGGACGAGGTGAACATCTGTACCGAC 2154
 DB 2101 CTGCAAGGCGATCAAGCTGCGCTGTGACAGAGCGGACGAGGTGAACATCTGTACCGAC 2160
 QY 2155 AGCCGATACGCGCTGGGCGATCATCCAGGCGCGCAAGGAGAGAGAGGAGCTGAGT 2214
 DB 2161 AGCCGATACGCGCTGGGCGATCATCCAGGCGCGCAAGGAGAGAGAGGAGCTGAGT 2220
 QY 2215 AACGAGATCATGAGAGAGCTGTATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGCGCGC 2274
 DB 2221 AACGAGATCATGAGAGAGCTGTATCAAGAGAGAGAGGTGTACTGTAGCTGGGTGCGCGC 2280
 QY 2275 CACAGAGGCGATCGCGGCAAGAGATGACAGAGCTGTGTAGAGAGAGGCGATCTCGCAG 2334
 DB 2281 CACAGAGGCGATCGCGGCAAGAGATGACAGAGCTGTGTAGAGAGAGGCGATCTCGCAG 2340
 QY 2335 GTGCTGTCTGTGAGAGCGCATGATGGCGGCGCATGTGATTAACAGATCAAGACGACCTG 2394
 DB 2341 GTGCTGTCTGTGAGAGCGCATGATGGCGGCGCATGTGATTAACAGATCAAGACGACCTG 2400
 QY 2395 TACGTGAGGAGCGCGCGCTTAGATTCATTAAGCTTCCGCGGCTTAGACCGGAGTA 2454
 DB 2401 TACGTGAGGAGCGCGCGCTTAGATTCATTAAGCTTCCGCGGCTTAGACCGGAGTA 2460
 QY 2455 TTC 2457
 DB 2461 TTC 2463

RESULT 3
 AX455914 2469 bp DNA linear PART 06-JUL-2002
 LOCUS AX455914 Sequence 30 from Patent WO0204493.
 DEFINITION AX455914
 ACCESSION AX455914
 VERSION AX455914.1 GI:21714899
 KEYWORDS
 ORGANISM
 SOURCE
 ORGANISM
 other sequences; artificial sequences.
 REFERENCE
 1 zur Megeide, J., Barnett, S.W., Engelbrecht, S. and van Rensburg, B.
 AUTHORS
 TITLE
 JOURNAL
 Polynucleotides encoding antigenic hiv type c polypeptides,
 Patent: WO 0204493-A 30 17-JAN-2002;
 CHIRON CORPORATION (US) ; University of Stellenbosch (ZA)

FEATURES
 source
 Location/Qualifiers
 1..2469
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="PR975(+)"

ORIGIN
 Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
 Beef Local Similarity 99.3%; Pred. No. 3.3e-199;
 Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;

QY 1 GTGACGCGCACCATGCGCGAGGCGCATGAGCGCACGAGCGCAAGATCTGTATGCGAG 60
 DB 1 GTGACGCGCACCATGCGCGAGGCGCATGAGCGCACGAGCGCAAGATCTGTATGCGAG 60
 QY 61 CGCAGCACTTCAAGAGGCGCCCAAGCGCATCATAGTGTCTCAACTGCGGCAAGAGGCGC 120
 DB 61 CGCAGCACTTCAAGAGGCGCCCAAGCGCATCATAGTGTCTCAACTGCGGCAAGAGGCGC 120
 QY 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAGAGTGGCGAAGAG 180
 DB 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAGAGTGGCGAAGAG 180
 QY 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAGAGTGGCGAAGAG 180
 DB 121 CACATGCGCGCGCAACTGCGCGCGCGCGCGCAAGAGGCGCTGTGAGAGTGGCGAAGAG 180
 QY 181 GGCACACAGATGAGAGAGCTGACACGAGCGCACGACCACTTCTTCCGCGAGAGACTGCGC 240
 DB 181 GGCACACAGATGAGAGAGCTGACACGAGCGCACGACCACTTCTTCCGCGAGAGACTGCGC 240
 QY 241 TTCCCCAGAGGAGGCGCGGAGTTCCCCAGCGAGACAGAACCGCGCCACAGCCCGCAC 300
 DB 241 TTCCCCAGAGGAGGCGCGGAGTTCCCCAGCGAGACAGAACCGCGCCACAGCCCGCAC 300
 QY 301 AGCGCGAGCTGAGAGGTGCGGCGGACCAACCGCGCGAGAGGCGGCGCGAGGCGCGAG 360
 DB 301 AGCGCGAGCTGAGAGGTGCGGCGGACCAACCGCGCGAGAGGCGGCGCGAGGCGCGAG 360
 QY 361 GGCACCTGAACTTCCCGCAAGATCACTGTGTGAGAGCGCGCGCTGTGAGATCAAGAGT 420
 DB 361 GGCACCTGAACTTCCCGCAAGATCACTGTGTGAGAGCGCGCGCTGTGAGATCAAGAGT 420
 QY 421 GCGCGCGAGATCAAGAGGCGCTGTGTGACACCGGCGCGGACGACACCGTGTGAGAGAG 480
 DB 421 GCGCGCGAGATCAAGAGGCGCTGTGTGACACCGGCGCGGACGACACCGTGTGAGAGAG 480
 QY 481 ATGAGCGTGGCGGAGAGTGAAGGCGCAAGTATGCGCGGATGCGGCGCTTCAATCAAG 540
 DB 481 ATGAGCGTGGCGGAGAGTGAAGGCGCAAGTATGCGCGGATGCGGCGCTTCAATCAAG 540
 QY 541 GTGCGCGAGTACGACCAATCTGTATGAGATCTTGTGCGGCAAGAGGCGATCGGCACTG 600
 DB 541 GTGCGCGAGTACGACCAATCTGTATGAGATCTTGTGCGGCAAGAGGCGATCGGCACTG 600
 QY 601 CTGATGCGCGCGCGCGCGGTGAACTCATCTGCGCGCAATCTGACCCAGCTGCGGCTGCG 660
 DB 601 CTGATGCGCGCGCGCGCGGTGAACTCATCTGCGCGCAATCTGACCCAGCTGCGGCTGCG 660
 QY 661 ACCCTGAACTTCCCGCATGAGCGCGCATGAGACCGTGTGAGAGCTGAAAGGCGGCGAGT 720
 DB 661 ACCCTGAACTTCCCGCATGAGCGCGCATGAGACCGTGTGAGAGCTGAAAGGCGGCGAGT 720
 QY 721 GACGCGCGCGCAAGTGAAGAGTGGCGCTGTGACCGAGAGAGATGAAGGCGCTGTGACGCGC 780
 DB 721 GACGCGCGCGCAAGTGAAGAGTGGCGCTGTGACCGAGAGAGATGAAGGCGCTGTGACGCGC 780
 QY 781 ATCTGCGAGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCTCGCGAGAACCTCTAC 840
 DB 781 ATCTGCGAGAGATGAGAGAGGCGCAAGATCACCAAGATCGGCTCGCGAGAACCTCTAC 840
 QY 841 AACACCGCGTGTGTGCGCATCAAGAGAGAGACACCAAGTGGCGGCAAGCTGTGTGAGC 900
 DB 841 AACACCGCGTGTGTGCGCATCAAGAGAGAGACACCAAGTGGCGGCAAGCTGTGTGAGC 900
 QY 901 TTCCGCGAGTGAACAGAGCGACCAAGACTTCTGTGAGAGTGAAGTGAAGTGAAGTGAAG 960

Db 901 TTCCGGAAGCTGAAACAGGCGACCCAGAGATTCTGGAGGTGAGCTGGGATCTCCCGAC 960
Qy 961 CCCGCGGCTGAAAGAAAGAGAGCGTGAACGTCGTGAACTGGGCGAAGCGCTTACTTC 1020
Db 961 CCCGCGGCTGAAAGAAAGAGAGCGTGAACGTCGTGAACTGGGCGAAGCGCTTACTTC 1020
Qy 1021 AGCTGCCCCCTGAGCAGAGACTTCGCGAAGTACACCGCTTTCACCTCCAGCATCAAC 1080
Db 1021 AGCTGCCCCCTGAGCAGAGACTTCGCGAAGTACACCGCTTTCACCTCCAGCATCAAC 1080
Qy 1081 AACGAGACCCCGGCAATCCGCTACAGTCAACATGTGCGCCAGGAGCTGAAGGGCAGC 1140
Db 1081 AACGAGACCCCGGCAATCCGCTACAGTCAACATGTGCGCCAGGAGCTGAAGGGCAGC 1140
Qy 1141 CCCAGCATCTTCAAGAGCAGCATGACCAAGATCTTGAAGCCCTTCGCGCGCCGCAACCC 1200
Db 1141 CCCAGCATCTTCAAGAGCAGCATGACCAAGATCTTGAAGCCCTTCGCGCGCCGCAACCC 1200
Qy 1201 GAGATGTGATCTTACCA-----GGCCCCCTGTACGTGGGCGAGCACTTGAGATCGGC 1254
Db 1201 GAGATGTGATCTTACCA-----GGCCCCCTGTACGTGGGCGAGCACTTGAGATCGGC 1260
Qy 1255 CAGCAACGCGCCAAAGTTCAGAGAGCTGCGGAGACCTGTGCGTGGGCTTACACACC 1314
Db 1261 CAGCAACGCGCCAAAGTTCAGAGAGCTGCGGAGACCTGTGCGTGGGCTTACACACC 1320
Qy 1315 CCCGCAAGAGAGCAGCAAG 1368
Db 1321 CCCGCAAGAGAGCAGCAAG 1380
Qy 1369 GACAGATGAGACCTGTGAG 1428
Db 1381 GACAGATGAGACCTGTGAG 1440
Qy 1429 ATCCAGAGCTGTGGGCAAGCTGGAATCTGGGCGAGCAGATCTAACCCCGGCAAGAGTG 1488
Db 1441 ATCCAGAGCTGTGGGCAAGCTGGAATCTGGGCGAGCAGATCTAACCCCGGCAAGAGTG 1500
Qy 1489 CGCGCACTGTGCAAGCTGTGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1548
Db 1501 CGCGCACTGTGCAAGCTGTGCGCGCGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Qy 1549 GAGAGAGCGGAGCTGAGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1608
Db 1561 GAGAGAGCGGAGCTGAGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Qy 1609 GTGTACTAGAGACCCCGAG 1668
Db 1621 GTGTACTAGAGACCCCGAG 1680
Qy 1669 TGGACCTTACAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1728
Db 1681 TGGACCTTACAGATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy 1729 ATGCGAGCGGCGCAACCAAG 1788
Db 1741 ATGCGAGCGGCGCAACCAAG 1800
Qy 1789 ATGAGAGAGATCTGTGATCTGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1848
Db 1801 ATGAGAGAGATCTGTGATCTGTGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy 1849 ACCTGGAGAGACTGTGAG 1908
Db 1861 ACCTGGAGAGACTGTGAG 1920
Qy 1909 GTGAAACACCCCGCTGTGAG 1968
Db 1921 GTGAAACACCCCGCTGTGAG 1980
Qy 1969 GCGGAGACTTCTTACGTGAG 2028

Db 1981 GCGAGAGCTTCTACGTGAG 2040
Qy 2029 TACGTGAG 2088
Db 2041 TACGTGAG 2100
Qy 2089 ACCGAGCTGAG 2148
Db 2101 ACCGAGCTGAG 2160
Qy 2149 ACCGAG 2208
Db 2161 ACCGAG 2220
Qy 2209 CTGTGAG 2268
Db 2221 CTGTGAG 2280
Qy 2269 CCGGCGCAAG 2328
Db 2281 CCGGCGCAAG 2340
Qy 2329 CCGAGAGCTGTGATCTGTGAG 2388
Db 2341 CCGAGAGCTGTGATCTGTGAG 2400
Qy 2389 GACCTGTACGTGGGAG 2448
Db 2401 GACCTGTACGTGGGAG 2460
Qy 2449 GGTGAATTC 2457
Db 2461 GGTGAATTC 2469

RESULT 4
BD263705
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
OS
PM
PD
PR
PI
PC
Key
FT
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

ORIGIN

Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best Local Similarity 93.6%; Pred. No. 6.3e-167;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

QY 170 GCGGCAAGAGGCGCCACAGATGAAGACTGCAACGAGCGCCAGGCCAATTCTTCCGCG 229
DB 1 GCGGCGCGAAGAGCACCAATGAAGATTGCACTGAGAGACAGGCTAATTTCTTCCGCG 60
QY 230 AGGACTTGCCCTTCCCCCGAGGCAAGGCCCGGAGATTCCCCAGCGAGCAACCGCGCA 289
DB 61 AGGACTTGCCCTTCTTGCAAGGCGACGCGCGAGTTCAAGCGACGACCGCGCA 120
QY 290 ACAAGCCCAACGACCGCGAGCTGAGGTGCGGCGG-----ACAACCCCGCGAGG 343
DB 121 ACAAGCCCAACGCGCGAGCTGAGGTGCGGCGGCGAGAACAAACGCTGAGGAGG 180
QY 344 CGGCGCGCCGAGCGCGAGGCAACCTG-----AATTCCCCAGATCAACCTGTGCAAG 397
DB 181 CGGCGCGCGACCGCGAGGCAACGTTGACTTCACTTCCCGAGATCAACCTGTGCAAG 240
QY 398 GCCCTCTGTGATCAAGAGTGGCGCGCAATCAAGAGGCCCTGTGCAACCGCG 457
DB 241 GCCCTCTGTGATCAATCAAGATCGCGCGCAAGTCAAGAGGCCGCTGCAACCGCG 300
QY 458 CGGACGACCGGTGAGAGATGAGGCTGCCGCGCAAGTGAAGCCCAAGATGATG 517
DB 301 CGGACGACCGGTGAGAGATGAGCTGCCGCGCAAGTGAAGGCCCAAGATGATG 360
QY 518 GCGGACCGCGCGCTTCAATCAAGTGGCGCAAGTACGACCAAGATCTGATCGAGATCTGCG 577
DB 361 GCGGAGTCGGGGGCTTCAATCAAGTGGCGCAAGTACGACCAAGATCTGATCTGCG 420
QY 578 GCGAAGAGGCTATCGGCAACGTTGCTGATGCGCCCCCAACCTGTGAACATCATCGCGCA 637
DB 421 GCGAAGAGGCTATCGGCAACGTTGCTGATGCGCCCCCAACCTGTGAACATCATCGCGCA 480
QY 638 ACATGCGGCGCGAGCTGAGCTGCAACCTGAACCTTCCCGCATGAGCCCATCGAGACGTTGC 697
DB 481 ACCTGCTGACCGAGATCGGCTGACCTTGAACCTTCCCGCATGAGCCCATCGAGACGTTGC 540
QY 698 CGGTGAAGCTGAAGCCCGGCAATGACGAGCCCAAGGTGAAGCATGAGCCCTGACCGAG 757
DB 541 CGGTGAAGCTGAAGCCCGGCAATGAGCGGCCCAAGGTGAAGCATGAGCCCTGACCGAG 600
QY 758 AGAAGATCAAGGCCCTGACCGCGCATTTGCGAGAGATGGAAGAGAGGCCAAGATCAACA 817
DB 601 AGAAGATCAAGGCCCTGAGTGAATCTGACCGAGATGGAAGAGAGGCCAAGATCAACA 660
QY 818 AGATGCGGCGCGAGAACCCCTTCAACACCCCGTGTGGCATTCAGAGAGAGAGCAACA 877
DB 661 AGATGCGGCGCGAGAACCCCTTCAACACCCCGTGTGGCATTCAGAGAGAGAGCAACA 720
QY 878 CCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGCG 937
DB 721 CCAAGTGGCGCAAGCTGTGACTTCCGCGAGCTGAACAAGCGCACCCAGACTTCTGCG 780
QY 938 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTGAAGAGAGAGAGAGAGAGCGTGC 997
DB 781 AGGTGAGCTGGGCAATCCCCCAACCCCGCGGCTGAAGAGAGAGAGAGAGAGCGTGC 840
QY 998 TGAAGGTGGGCGACCGCTTCAAGGTGCGCTTGAACGAGACTTCCGCAAGTACACCG 1057
DB 841 TGAAGGTGGGCGACCGCTTCAAGGTGCGCTTGAACGAGACTTCCGCAAGTACACCG 900
QY 1058 CCTTCAACATCCCGAGCATCAACAAGAACCCCGGCGATCCGCTACAGTCAAGTGC 1117
DB 901 CCTTCAACATCCCGAGCATCAACAAGAACCCCGGCGATCCGCTACAGTCAAGTGC 960
QY 1118 TGCCTCCAGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGAGAGATGACCAAGATCTGC 1177
DB 961 TGCCTCCAGGCTGGAAGGCGAGCCCGACGATCTTCCAGAGAGAGATGACCAAGATCTGC 1020

QY 1178 AGCCCTTCCGCGCCCGCAACCCCGAGATCGTATCTACCAAGGCCCCCTGTACGTGGCA 1237
DB 1021 AGCCCTTCCGCGCAAGAGAGAACCCCGACATCGTATCTTACCAAGGCCCCCTGTACGTGGCA 1080
QY 1238 GCGACTTGAGATCGGCGACGACCGCGCCAGAGTGAAGAGCTGCGCAAGCATCTGTGC 1297
DB 1081 GCGACTTGAGATCGGCGACGACCGCACCAAGATGAAGAGCTGCGCGACCATGTGC 1140
QY 1298 GCTGGGCGCTTCAACACCCCGCAAGAGAGCCCAAGAGGCCCTTCCGTGCCATG 1357
DB 1141 GCTGGGCGCTTCAACACCCCGCAAGAGAGCCCAAGAGAGGCCCTTCCGTGCCATG 1200
QY 1358 AGCTGACCCCGCAACAAGTGAACCGTGCAGGCCCATCGAGCTGCGCGAGAGAGAGACTGGA 1417
DB 1201 AGCTGACCCCGCAACAAGTGAACCGTGCAGGCCCATCGAGCTGCGCGAGAGAGACTGGA 1260
QY 1418 CGGTGAAGCAATCCAGAAAGCTGTGGCAAGCTGAATGGGCGACGACGATCTACCCCG 1477
DB 1261 CGGTGAAGCAATCCAGAAAGCTGTGGCAAGCTGAATGGGCGACGACGATCTACCCCG 1320
QY 1478 GCATCAAGTGGCGGACGTTGTGCAAGCTGTGCGCGCGCCAGAGCCCTGACCGACATG 1537
DB 1321 GCATCAAGTGAAGGAGCTGTGCAAGCTGTGCGCGCGCACCAAGGCCCTGACCGAGTGA 1380
QY 1538 TGCCTCTGACCGAGGCGGAGCTGTGAGCTGGCGAGAACCGCGAGATCTTGCAGG 1597
DB 1381 TGCCTCTGACCGAGGCGGAGCTGTGAGCTGGCGAGAACCGCGAGATCTTGAAGAG 1440
QY 1598 CGGTGACGCGCTGTACTTACGACCCCGACAGAACCTTGTGGCGAGATTCAGAAAGCAG 1657
DB 1441 CGGTGACGAGGTGTACTTACGACCCCGACAGAACCTTGTGGCGAGATTCAGAAAGCAG 1500
QY 1658 GCGACGACCAAGTGAACCTTACCAAGATCTTACAGAGGCCCTTGAAGAACCTTGAAGACCGGCA 1717
DB 1501 GCGAGGCGCAAGTGAACCTTACCAAGATCTTACAGAGGCCCTTGAAGAACCTTGAAGACCGGCA 1560
QY 1718 AGTACGCGCAAGATGCGACCGCGCCACCAACGAGCTGAAGAGAGCTGACCGAGGCGTGC 1777
DB 1561 AGTACGCGCGATGCGCGCGCGCCACCAACGAGCTGAAGAGAGCTGACCGAGGCGTGC 1620
QY 1778 AGAAGATCGCCATGAGAGCATGTGATCTGGGGCAAGACCCCGAAGTTCCGCTGCCCA 1837
DB 1621 AGAAGTGAAGCACGAGAGCATGTGATCTGGGGCAAGATTCGCAAGTTCAAGCTGCCCA 1680
QY 1838 TTCAGAGAGACCTTGGAGAGACTGTGACCGACTTACCTGAGAGGCGACCTGAGATCCCG 1897
DB 1681 TTCAGAGAGACCTTGGAGAGGCTGTGATGAGAGTACCTGAGAGGCGACCTGAGATCCCG 1740
QY 1898 AGTGGAGTGTGTAACACCCCGCGTGTGAACCTGTGTAACAGCTGAGAGAGAGG 1957
DB 1741 AGTGGAGTGTGTAACACCCCGCGTGTGTAACCTGTGTAACAGCTGAGAGAGAGG 1800
QY 1958 CCATCATCGCGCGCGACCTTCTTACGTGACGCGCGCGCAACCGCGAGACCAAGATG 2017
DB 1801 CCATGTGGGCGCGAGACCTTCTTACGTGAGCGCGCGCAACCGCGAGACCAAGATG 1860
QY 2018 GCAAGCGCGCTTACGTGACCGACCGGGCGCGCAAGAGTGTGAGCTGACCGAGACCA 2077
DB 1861 GCAAGCGCGCTTACGTGACCGACCGGGCGCGCAAGAGTGTGAGCTGACCGAGACCA 1920
QY 2078 CCAACGAGAGCGAGGCTGAGGCGCATCAGCTGAGCCCTGAGAGACAGCGGCAAGAG 2137
DB 1921 CCAACGAGAGCGAGGCTGAGGCGCATCAGCTGAGCCCTGAGAGACAGCGGCTGAGAG 1980
QY 2138 TGAACATGTGACCGACAGCGCTGAGCGCTTGGGATCATCCAGGCCACCGCGACAG 2197
DB 1981 TGAACATGTGACCGACAGCGCTGAGCGCTTGGGATCATCCAGGCCACCGCGACAG 2040
QY 2198 GCGAGAGGAGCTGTGTAACCAAGATCATGAGAGCTGATCAAGAGAGAGAGTGTACC 2257
DB 2041 GCGAGAGGAGCTGTGAGCGAGATCATGAGAGCTGATCAAGAGAGAGAGTGTACC 2100

1558 GCCACGACGAGTGAAGCTTACAGAGAGCCCTTCAAGAACTGAAAGCCGCA 1717
1501 GCCAGGCGCAGTGAAGCTTACAGAGAGCCCTTCAAGAACTGAAAGCCGCA 1560
1718 AGTACGCGCAAGATGCGACCGCCCAACCAAGACGTGAAGCTGAACGAGCCGTGC 1777
1561 AGTACGCGCGCAGTGAAGCGCCGCGCCACCAAGACGTGAAGCGCTGAACGAGCCGTGC 1620
1778 AGAAGATCGCCATGAGAGCATCTGTGATCTTGGGGCAAGACCCCCCAAGTTCCGCTGCCCA 1837
1621 AAGAGGTGAGCACCGAGAGCATCTGTGATCTTGGGGCAAGATCCCAAGTTCAAGCTGCCCA 1680
1838 TCAGAGAGAGAGCTTGGAGAGCTTGGTGAACCGACTGAGAGCCCACTTGGATCCCG 1897
1681 TCAGAGAGAGAGCTTGGAGAGCTTGGTGAACCGACTGAGAGCCCACTTGGATCCCG 1740
1898 AGTGGAGTTCTGTGAACACCCCTCTGTGTGAAGCTGTGTGAACAGCTGAGAGAGAGC 1957
1741 AGTGGAGTTCTGTGAACACCCCTCTGTGTGAAGCTGTGTGAACAGCTGAGAGAGAGC 1800
1958 CGATCATGCGCGCCGAGACCTTCTAAGTGAAGCGCGCCGCAACCGCGAGACCAAGATCG 2017
1801 CGATGTTGGCGCGCGAGACCTTCTAAGTGAAGCGCGCCGCAACCGCGAGACCAAGCTGC 1860
2018 GCAAGGCGCGCTACGTCAGACCGAGCGCGCGAGAGATCGTGAAGCTGAACCGAGACCA 2077
1861 GCAAGGCGCGCTACGTCAGACCGAGCGCGCGAGAGATCGTGAAGCTGAACCGAGACCA 1920
2078 CCAACCGAGAGACCGAGCTGAGAGCCATCAAGCTGAGCCCTGAGAGACAGCGCGAGAGC 2137
1921 CCAACCGAGAGACCGAGCTGAGAGCCATCAAGCTGAGCCCTGAGAGACAGCGCGAGAGC 1980
2138 TGAATCATGTCAGACCGAGCTGAGAGCCCTGAGAGATCAATCAAGGCGCCGAGAGACCA 2197
1981 TGAATCATGTCAGACCGAGCTGAGAGCCCTGAGAGATCAATCAAGGCGCCGAGAGACCA 2040
2198 GCGAGAGCGAGCTGAGAGACCGATCAATCAAGAGCTGATCAAGAGAGAGAGAGTGTACC 2257
2041 GCGAGAGCGAGCTGAGAGACCGATCAATCAAGAGCTGATCAAGAGAGAGAGAGTGTACC 2100
2258 TGAAGTGTGTCGCGCCCAAGAGGATGAGCGGCAACGAGAGATCGAACAGCTGTGTA 2317
2101 TGGCGCTGGGTGCGCCGCCCAAGAGGATGAGCGGCAACGAGAGATCGAACAGCTGTGTA 2160
2318 GGAAGGCGATCGCGAAGGTCGTCTTGGAGCGGATCGAATGGCGCATCTGATCTTACC 2377
2161 GCGCGGCGATCGCGAAGGTCGTCTTGGAGCGGATCGAATGGCGCATCTGATCTTACC 2220
2378 AGTACATGAGACGCTGTACGTGGGCGAGCGGCGCCCTAGAGATCGATTAAAGCTTCCCG 2437
2221 AGTACATGAGACGCTGTACGTGGGCGAGCGGCGCCCTAGAGATCGATTAAAGCTTCCCG 2280
2438 GGGCTAGACCGCGTGAATTC 2457
2281 GGGCTAGACCGCGTGAATTC 2300

RESULT 6
AR373388 2300 bp DNA linear PAT 18-DEC-2003
LOCUS AR373388
DEFINITION Sequence 83 from patent US 6602705.
ACCESSION AR373388
VERSION AR373388.1 GI:40075491
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2300)
AUTHORS Barnett,S.W., Mesede,J.J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 83 05-AUG-2003;

Chiron Corporation; Emeryville, CA
FEATURES
source
1..2300
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN
Query Match 83.0%; Score 2040; DB 6; Length 2300;
Best Local Similarity 93.6%; Pred. No. 6.3e-167;
Matches 2153; Conservative 0; Mismatches 135; Indels 12; Gaps 2;

170 GCGGCAAGAGAGGCGCCACCAATGAAGACTGACCGAGCGCCAGGCAATCTTCTCGCG 229
1 GCGGCGGGAAGAGACCAATGAAGATTCACATGAGAGACAGGCTAATTTCTTCCCG 60
230 AGGACTTGGCTTCCCCCGAGGCGAAGGCGCGAGTTCCCCAGAGAGAGAACCGCGCA 289
61 AGGACTTGGCTTCTTGAAGGCGAAGGCGCGAGTTCAAGAGAGAGAGACCGCGCA 120
290 AAGGCGCCAGCGCGAGCTGACAGTGCAGCGCGG-----ACAAACCCCGCAGCGAG 343
121 AAGGCGCCAGCGCGCGAGCTGACAGTGCAGCGCGG-----ACAAACCCCGCAGCGAG 180
344 CCGGCGCGAGCGCGCGAGCGCACTCTG-----AATTCCCGCAGATCACTCTGTGAGC 397
181 CCGGCGCGAGCGCGCGAGCGCACTCTGAGCTTCAACTTCCCGCAGATCACTCTGTGAGC 240
398 GCCCTCTGTGAGATCAAGAGTGGCGCGCGAGATCAAGAGAGCGCTTGTGAGACCGCG 457
241 GCCCTCTGTGAGATCAAGAGTGGCGCGCGAGATCAAGAGAGCGCTTGTGAGACCGCG 300
458 CCGAGACCAACCGTGTGAGAGAGATGAGCTGCGCGCGAGATGAGAGCGCCAAAGATATCG 517
301 CCGAGACCAACCGTGTGAGAGAGATGAGCTGCGCGCGAGATGAGAGCGCCAAAGATATCG 360
518 GCGGATGCGGCGCTTCAATCAAGTGCAGAGTACAGACAGATCTCTGATCGAGATCTGCG 577
361 GCGGATGCGGCGCGCTTCAATCAAGTGCAGAGTACAGACAGATCTCTGATCGAGATCTGCG 420
578 GCAAGAGGCGATCGGACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 637
421 GCGAGAGGCGATCGGACCGTGTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 480
638 ACATCTGACCGAGCTGAGAGCGCTGAGAGCTTCCCGCATGAGCGCCCATGAGAGCGTGC 697
481 ACCTGTGACCGAGATCGGCTGAGAGCGCTGAGAGCTTCCCGCATGAGCGCCCATGAGAGCGTGC 540
698 CCGTGAAGCTGAAGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 757
541 CCGTGAAGCTGAAGCGCGCGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
758 AGAAGATCAAGCGCTTGAAGCGCGCATCTGCGAGAGAGATGAGAGAGAGAGAGATCAACA 817
601 AGAAGATCAAGCGCGCTTGAAGCGCGCATCTGCGAGAGAGATGAGAGAGAGAGAGATCAACA 660
818 AGATGCGCGCGAGAGACCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 877
661 AGATGCGCGCGAGAGACCGCTTGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
878 CCAAGTGAAGAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCGCACTTCTGCG 937
721 CCAAGTGAAGAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCGCACTTCTGCG 780
938 AGTGAAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCGCACTTCTGCG 997
781 AGTGAAGCTGTGAGCTTTCGCGAGAGCTGAACAGAGCGCACTTCTGCG 840
998 TGAAGTGTGAGAGCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1057
841 TGAAGTGTGAGAGCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
1058 CTTTCAACCATCCCGAGATCAACAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1117

Db CCTTCAACCATCCCGACATCAACAAGAACCCCGGATCCGGTACCAATCAACATGTC 960
 QY 1118 TCCCCAGGGCTGGAAGGGGAGCCCGAGCATCTTTCAGAGCAGATGACCAAGATCTGG 1177
 Db 961 TCCCCAGGGCTGGAAGGGGAGCCCGGATCTTTCAGAGCAGATGACCAAGATCTGG 1020
 QY 1178 AGCCCTTCGGCCCGCCGCAACCCCGAGATGATGATCTCAACAGGCCCCCTGTACGTGGGA 1237
 Db 1021 AGCCCTTCGGCCCGCAAGAACCCCGAGATGATGATCTCAACAGGCCCCCTGTACGTGGGA 1080
 QY 1238 GGAACCTGAGATTCGGCCAGCAACCGCCAGATTCAGAGAGCTGCGCAACACTGCTGC 1297
 Db 1081 GGAACCTGAGATTCGGCCAGCAACCGCAACAGATTCAGAGAGCTGCGCCAGCAACACTGCTGC 1140
 QY 1298 GCTGGGGCTTCAACACCCCGCAAGAGACACAGAGAGAGCCCTTCTGCTCCCATG 1357
 Db 1141 GCTGGGGCTTCAACACCCCGCAAGAGACACAGAGAGAGCCCTTCTGCTCCCATG 1200
 QY 1358 AGCTGACCCCGCAAGATGAGACCTGTGACAGCCCATGAGCTGCGCGAGAGAGAGATGGA 1417
 Db 1201 AGCTGACCCCGCAAGATGAGACCTGTGACAGCCCATGAGCTGCGCGAGAGAGAGATGGA 1260
 QY 1418 CCGTGAACGACATCCAGAGCTGTGAGGAGCTGAACTGGGCGACGACGATCTACCCCG 1477
 Db 1261 CCGTGAACGACATCCAGAGCTGTGAGGAGCTGAACTGGGCGACGACGATCTACCCCG 1320
 QY 1478 GCATCAAGGTGCGCCAGCTGTGAGAGCTGTGCGCGCGCCAGAGCCCTGACCGACATG 1537
 Db 1321 GCATCAAGGTGAGAGAGCTGTGAGAGCTGTGCGCGCGCCAGAGCCCTGACCGAGTGA 1380
 QY 1538 TGGCCCTGACCGAGAGGCGCGAGCTGTGAGAGCTGTGCGCGCGCCAGAGCTGTGCGGAGC 1597
 Db 1381 TGGCCCTGACCGAGAGGCGCGAGCTGTGAGAGCTGTGCGCGCGCCAGAGCTGTGAGAGC 1440
 QY 1598 CCGTGACCGGCGGTACTGACACCCAGAGAGAGCTGTGCGCGCGAGATCTGAGAGCAG 1657
 Db 1441 CCGTGACCGAGGTACTGACACCCAGAGAGAGCTGTGCGCGCGAGATCTGAGAGCAG 1500
 QY 1658 GCCACAGCAGTGAAGCTTACAGATCTACAGAGAGCCCTTCAAGAACCTGAGACCGGCA 1717
 Db 1501 GCCACAGCAGTGAAGCTTACAGATCTACAGAGAGCCCTTCAAGAACCTGAGACCGGCA 1560
 QY 1718 AGTACGCCAAGATGCGGACCGCCCAACCAAGACCTGAAAGCAGCTGACCGAGCCGTGC 1777
 Db 1561 AGTACGCCCGCATGCGCGCGCCCAACCAAGACCTGAAAGCAGCTGACCGAGCCGTGC 1620
 QY 1778 AGAAGATCCGATGAGAGAGATGATCTGTGGGCAAGACCCCAAGTTCCGCTGCCCA 1837
 Db 1621 AGAAGATGAGACCGAGAGATGATCTGTGGGCAAGATCCCAAGTTCCAGCTGCCCA 1680
 QY 1838 TCCAGAGAGAGACCTGAGAGACCTGTGAGACCACTACTGAGAGGCACTGTGATCCCG 1897
 Db 1681 TCCAGAGAGAGACCTGAGAGAGCTGTGAGATGAGATGAGAGGCACTGTGATCCCG 1740
 QY 1898 AGTGGAGTTCTGGAACACCCCGCTGTGTAAGCTGTGTACCAAGCTGAGAGAGAGC 1957
 Db 1741 AGTGGAGTTCTGGAACACCCCGCTGTGTAAGCTGTGTACCAAGCTGAGAGAGAGC 1800
 QY 1958 CCAATATCGGCGCGAGACCTTCTAAGTGAACGCGCCCGCAACCGGAGACCAAGATG 2017
 Db 1801 CCAATCGGCGCGAGACCTTCTAAGTGAACGCGCGCGCAACCGGAGACCAAGCTG 1860
 QY 2018 GCAAGCGGCTAGTGAACCGAGCCGAGGAGAGAGATCTGAGAGCTGAGACCAAGACCA 2077
 Db 1861 GCAAGCGGCTAGTGAACCGAGCCGAGGAGAGAGATCTGAGAGCTGAGACCAAGACCA 1920
 QY 2078 CCAACCAAGAGACCGAGCTGAGAGCCATCAAGCTGCGCTTGCAGAGACGCGGACGAG 2137
 Db 1921 CCAACCAAGAGACCGAGCTGAGAGCCATCAAGCTGCGCTTGCAGAGACGCGGCTGAG 1980
 QY 2138 TGAACATCTGTGACCGAGACGATGAGCCCTGTGGCAATCAAGGCGCCAGCCGAGCAAG 2197
 Db 1981 TGAACATCTGTGACCGAGACGATGAGCCCTGTGGCAATCAAGGCGCCAGCCGAGCAAG 2040

QY 2198 GCGAGAGGAGCTGTGTAACAGATCATGAGAGCTGATGTAAGAGAGAGAGGNTACC 2257
 Db 2041 GCGAGAGGAGCTGTGTAACAGATCATGAGAGCTGATGTAAGAGAGAGAGGNTACC 2100
 QY 2258 TGAGCTGGGTGCGCCCGCAAGAGGATCGCGGCAAGAGAGATGCAAGAGCTGTGA 2317
 Db 2101 TGAGCTGGGTGCGCCCGCAAGAGGATCGCGGCAAGAGAGATGCAAGAGCTGTGA 2160
 QY 2318 GCAAGGCAATCCGAGAGGTGCTGTTCTGAGACGCGATGATGAGCGGATCTGATAC 2377
 Db 2161 GCGCGGATCGCAAGAGGTGCTGTTCTGAGACGCGATGATGAGCGGATCTGATAC 2220
 QY 2378 AGTACATGAGACCTGTACAGTGGGCAAGCGCGGCTGAGATGATGATGATGATG 2437
 Db 2221 AGTACATGAGACCTGTACAGTGGGCAAGCGCGGCTGAGATGATGATGATGATG 2280
 QY 2438 GGGCTAGCACCGGTGATTC 2457
 Db 2281 GGGCTAGCACCGGTGATTC 2300
 RESULT 7
 BD263704
 LOCUS
 DEFINITION Improved expression of HIV polypeptides and production of
 virus-like particles.
 ACCESSION BD263704
 VERSION BD263704.1 GI:33073472
 KEYWORDS JP 200253124-A/71.
 SOURCE
 ORGANISM
 synthetic construct
 other sequences; artificial sequences.
 REFERENCE
 1 (bases 1 to 2306)
 AUTHORS Barnett,S., Megede,J.Z., Srivastava,I., Lian,Y., Hartog,K., Liu,H.,
 Greer,C., Selby,M. and Walker,C.
 TITLE Improved expression of HIV polypeptides and production of
 virus-like particles
 JOURNAL
 Patent: JP 200253124-A 71 08-OCT-2002;
 CHIRON CORP
 COMMENT
 OS Artificial Sequence
 PN JP 200253124-A/71
 PD 08-OCT-2002 JP 2000591193
 PE 30-DEC-1999 US 60/114495, 01-DEC-1999 US 60/164471 PI
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/164471 PI
 SUSAN BARNETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI
 KARIN HARTOG,
 PI HONG LIU CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER PC
 C12N15/09,A61K31/711,A61K38/00,A61K48/00,A61P31/18,A61P37/02, PC
 C12N5/10,
 PC C12N7/00,C12P21/02,C12N15/00,C12N5/00,A61K37/02 CC
 Description of Artificial Sequence: FS(-).protonod.RTopt.YM FH Key
 FEATURES
 source
 FT
 Location/Qualifiers
 1..2306
 /organism="Artificial Sequence".
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 ORIGIN
 Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
 Best Local Similarity 93.2%; Pred. No. 3.9e-165;
 Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;
 QY 170 GCGGCAAGAGGCGCCACGATGAGAGATGCAACGAGCGGCAAGCTTCTTCCGGG 229
 Db 1 GCGGCGCGCAAGAGACCAAGATGAGATGCACTGAGAGAGAGGCTTAATTTCTTCCGG 60
 QY 230 AGGACCTGAGCTTCTTCCCGAGGAGAGGCGCGAGATTTCCCGAGAGAGAGAGAGAG 289
 Db 61 AGGACCTGAGCTTCTTCCCGAGGAGAGGCGCGAGATTTCAAGAGAGAGAGAGAGAG 120

QY 290 AAGCCCAACAGCCGAGCTGACAGTGCAGCGG-----ACAACCCCGCAGCGAG 343
 DB 121 AAGCCCAACAGCCGAGCTGACAGTGCAGCGGAGAAACAAGCCGAGCGAG 180
 QY 344 CGGCGCGAGCGCAGCGGACCTCG-----AAGCTCCCGAGATCAAGCTGTGGAC 397
 DB 181 CGGCGCGAGCGCAGCGGACCTGTGAGCTTCACTCCCGAGATCAAGCTGTGGAC 240
 QY 398 GCGCCCTGTGAGATCAAGGTGGCGGACAGATCAAGAGGCGCTGTGACAGCGCG 457
 DB 241 GCGCCCTGTGAGATCAAGGTGGCGGACAGCTCAAGAGGCGCTGTGACAGCGCG 300
 QY 458 CGGAGAGACCTGTGAGAGAGATGAGCTGCGCGCAGAGTGAAGCCAGAGATGAG 517
 DB 301 CGGAGAGACCTGTGAGAGAGATGAGCTGCGCGCAGAGTGAAGCCAGAGATGAG 360
 QY 518 GCGGAGTGGCGGCTTCAATCAAGGTGGCGGAGTCAAGATCTGTGATCGAGATCTGC 577
 DB 361 GCGGAGTGGCGGCTTCAATCAAGGTGGCGGAGTCAAGATCTGTGAGATCTGC 420
 QY 578 GGAAGAGGCAATGGGACCGGTGATGGGCGGACCGCGGTGAACATCGGCGGCA 637
 DB 421 GCAAGAGGCAATGGGACCGGTGATGGGCGGACCGCGGTGAACATCGGCGGCA 480
 QY 638 ACATGCTGACCGAGCTGGCTGACACCTGGAATTCGCCATCAGCCCATCGAGACCTGAG 697
 DB 481 ACCTGCTGACCGAGTGGCTGACACCTGGAATTCGCCATCAGCCCATCGAGACCTGAG 540
 QY 698 CGGTGAAGTGAAGCCCGGACATGAGACCGCGCGCAGAGTGAAGAGTGGCGCTGACCGAG 757
 DB 541 CGGTGAAGTGAAGCCCGGACATGAGACCGCGCGCAGAGTGAAGAGTGGCGCTGACCGAG 600
 QY 758 AGAAGATCAAGGCGCTGACCGGATCTGTGAGAGATGAGAGAGGCGCAGAGTCAACA 817
 DB 601 AGAAGATCAAGGCGCTGTGTGAGATCTGTGACCGAGATGAGAGAGGCGCAGAGTCAACA 660
 QY 818 AGATGAGCGCCCGGAGAACCTCTACAACACCCCGGTGTTCCGCAATCAAGAGAGAGACAGA 877
 DB 661 AGATGAGCGCCCGGAGAACCTCTACAACACCCCGGTGTTCCGCAATCAAGAGAGAGACAGA 720
 QY 878 CCAAGTGGCGCAGCTGTGAGATCTTCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGG 937
 DB 721 CCAAGTGGCGCAGCTGTGAGATCTTCGCGAGCTGAACAAGCGCACCCAGAGACTTCTGG 780
 QY 938 AGGTGACGCTGGGCAATCCCGCAACCCCGCGGCTGAAGAGAGAGCGGTGACCGTGC 997
 DB 781 AGGTGACGCTGGGCAATCCCGCAACCCCGCGGCTGAAGAGAGAGCGGTGACCGTGC 840
 QY 998 TGAAGTGGCGCAGCTGATCACTTCAAGCTGCGCTGTGACAGAGACTTCCGCAAGTACAG 1057
 DB 841 TGAAGTGGCGCAGCTGATCACTTCAAGCTGCGCTGTGACAGAGACTTCCGCAAGTACAG 900
 QY 1058 CCTTCAACATCCCGCAGATCAACAAGAGACCCCGGCAATCCGCTACCAAGTCAAGCTGC 1117
 DB 901 CCTTCAACATCCCGCAGATCAACAAGAGACCCCGGCAATCCGCTACCAAGTCAAGCTGC 960
 QY 1118 TGCAGGAGGCTGAAGGCGCAGCGGACGATCTTCAAGAGAGAGTGAACCAAGATCTGC 1177
 DB 961 TGCAGGAGGCTGAAGGCGCAGCGGACGATCTTCAAGAGAGAGTGAACCAAGATCTGC 1020
 QY 1178 AGCCCTTCCGCGCGGCAACCCCGAGATGTGATCTACAGAGGCGCCCTGTAGTGGGCA 1237
 DB 1021 AGCCCTTCCGCGCGGCAACCCCGAGATGTGATCTACAGAGGCGCCCTGTAGTGGGCA 1080
 QY 1238 GCGAGCTGAGATCGGCGCAGACCGCGCAGAGTCAAGAGTCAAGAGCTGCGCAAGACCTGCTGC 1297
 DB 1081 GCGAGCTGAGATCGGCGCAGACCGCGCAGAGTCAAGAGTCAAGAGCTGCGCAAGACCTGCTGC 1140
 QY 1298 GCTGGGCGCTTCAACACCCCGCAGAGAGACCAAGAGAGGCGCCCTTCTGTGGCAT-- 1355
 DB 1141 GCTGGGCGCTTCAACACCCCGCAGAGAGACCAAGAGAGGCGCCCTTCTGTGGATGG 1200

QY 1356 ----GAGCTGACCCCGCAGAGTGAACGTGACAGGCCCATCGAGCTGCCAGAGAGAGA 1411
 DB 1201 GCTAGAGCTGACCCCGCAGAGTGAACGTGACAGGCCCATCATGCTGTGCCAGAGAGAGA 1260
 QY 1412 GCTGAGCTGTGAAGACATTCAGAGAGCTGTGGGCAAGCTGAATCTGGGCGCAGCGATCT 1471
 DB 1261 GCTGAGCTGTGAAGACATTCAGAGAGCTGTGGGCAAGCTGAATCTGGGCGCAGCGATCT 1320
 QY 1472 ACCCGGCAATGAAGTGGCGCAGCTGTGACAGCTGTGCGGCGCAGAGGCGCTGACCG 1531
 DB 1321 ACCCGGCAATGAAGTGGAGAGCTGTGACAGCTGTGCGGCGCAGAGGCGCTGACCG 1380
 QY 1532 ACATGTGCGCCCTGACCGAGAGGCGCAGCTGAGCTGGCGAGAACCGCGAGATCTGTC 1591
 DB 1381 AGGTGATCCCTCTGACCGAGAGGCGCAGCTGTGAGCTGGCGAGAACCGCGAGATCTGTC 1440
 QY 1592 GCGAGCCGTGTGACAGGCTGTGATCTACGACCCCGCAGAGACCTGTGTGCGCAGATCTCAGA 1651
 DB 1441 AGGAGCCGTGTGACAGGCTGTGATCTACGACCCCGCAGAGACCTGTGTGCGCAGATCTCAGA 1500
 QY 1652 AGCAGGCGCAGACCAAGTGAACCTACAGATCTACAGAGGCGCTTCAAGAACCTGAAGA 1711
 DB 1501 AGCAGGCGCAGGCGCAGTGAACCTACAGATCTACAGAGGCGCTTCAAGAACCTGAAGA 1560
 QY 1712 CGGCAAGTACGCGCAGAGTGGCACCGGCGCACCAACCAAGAGTGAAGAGCTGACCGAG 1771
 DB 1561 CGGCAAGTACGCGCGCGCACCGGCGCGCGCACCAACCAAGAGTGAAGAGCTGACCGAG 1620
 QY 1772 CGGTGCAAGAAATGGCCATGAGAGACATGTGTGATCTGGGGCAGAACCCCGAAGTTCCGCG 1831
 DB 1621 CGGTGCAAGAAATGGCCATGAGAGACATGTGTGATCTGGGGCAGAAATCCCGAAGTTCAAGC 1680
 QY 1832 TGCCCATCAAGAGAGAGACTGTGAGACCTGTGAGACCGACTGAGAGCGACCTGGA 1891
 DB 1681 TGCCCATCAAGAGAGAGACTGTGAGACCTGTGAGAGCGCTGTGAGAGCGACCTGGA 1740
 QY 1892 TCCCGAGTGAAGTTCGTGAACACCCCGCTGTGTGAGAGCTGTGATCAAGCTGAGAG 1951
 DB 1741 TCCCGAGTGAAGTTCGTGAACACCCCGCTGTGTGAGAGCTGTGATCAAGCTGAGAG 1800
 QY 1952 AGGAGCCCATATGCGGCGCGCAGAGCTTCTACAGTGAACGCGCGCGCAGACCGCGAGACCA 2011
 DB 1801 AGGAGCCCATATGCGGCGCGCAGAGCTTCTACAGTGAACGCGCGCGCAGACCGCGAGACCA 1860
 QY 2012 AGATGCGCAGGCGGCTACGTGACCGGACCGGCGCGGCGCAGAGATCGTGAAGCTGACCG 2071
 DB 1861 AGATGCGCAGGCGGCTACGTGACCGGACCGGCGCGGCGCAGAGATCGTGAAGCTGACCG 1920
 QY 2072 AGACCAACCAACAGAGAGACCGAGCTGACAGGCAATCAGCTGAGCGCTGTGACAGACAGCGGCA 2131
 DB 1921 AGACCAACCAACAGAGAGACCGAGCTGACAGGCAATCAGCTGAGCGCTGTGACAGACAGCGGCG 1980
 QY 2132 GCGAGTGAACATGTGTGACCGACAGCGCAGTACCGCTGTGGGCAATCTCAAGGCGCGAGCGCG 2191
 DB 1981 TGAAGTGAACATGTGTGACCGACAGCGCAGTACCGCTGTGGGCAATCTCAAGGCGCGAGCGCG 2040
 QY 2192 ACAAGAGGAGAGGCGAGCTGTGAAACGAGATCATGAGCAGCTGATCAAGAGAGAGAGG 2251
 DB 2041 ACAAGAGGAGAGGCGAGCTGTGAAACGAGATCATGAGCAGCTGATCAAGAGAGAGAGG 2100
 QY 2252 TGTACTTGAAGTGTGGTGCCTCCCGCAAGAGGCAATCGCGGCGCAACGAGAGATTCAGCAAGC 2311
 DB 2101 TGTACTTGAAGTGTGGTGCCTCCCGCAAGAGGCAATCGCGGCGCAACGAGAGATTCAGCAAGC 2160
 QY 2312 TGTGTGAGAGAGGCAATCGCGCAAGTGTGTTCTGTGACGCGCATTCGATGGCGGCAATCTGTGA 2371
 DB 2161 TGTGTGAGAGAGGCAATCGCGCAAGTGTGTTCTGTGACGCGCATTCGATGGCGGCAATCTGTGA 2220
 QY 2372 TCTACAGATGATGAGAGACCTGTACGAGGAGAGCGGCGCGCTGTGAGATTCGATTAAGAGC 2431
 DB 2221 TCTACAGATGATGAGAGACCTGTACGAGGAGAGCGGCGCGCTGTGAGATTCGATTAAGAGC 2280
 QY 2432 TTCCCGGCGCTGACACCGGTGAATTC 2457

Db 2281 TTCCCGGGGCTAGCACCGGTGAATTC 2306

RESULT 8
LOCUS CQ870574 2306 bp DNA linear PAT 13-SRP-2004
DEFINITION Sequence 82 from Patent EP1433851.
ACCESSION CQ870574
VERSION CQ870574.1 GI:52000090
KEYWORDS
SOURCE
ORGANISM
synthetic construct
other sequences; artificial sequences.
REFERENCE
1
AUTHORS Barnett,S., Zumegege,J., Srilavastava,I., Lian,Y., Hartog,K.,
Liu,H., Greer,C., Selby,M. and Walker,C.
TITLE Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL Patent: EP 1433851-A 82 30-JUN-2004;
CHIRON CORPORATION (US)
FEATURES
source
1..2306
Location/Qualifiers
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="Description of Artificial Sequence:
FS (-).protopl.Rtopc.YM"

ORIGIN

Query Match 82.2%; Score 2019.2; DB 6; Length 2306;
Best Local Similarity 93.2%; Pred. No. 3..9e-165;
Matches 2150; Conservative 0; Mismatches 138; Indels 18; Gaps 3;

QY 170 GCGGCAAGAGGGGCAACGATGAGAGCTGACCGAGCGCCAGGCAACTTCTTCGCG 229
DB 1 GCGGCGCGAAGAGACCAATGAAAGATTGCACTGAGAGACAGGCTAATTTCTTCGCG 60

QY 230 AGGACTGTGCTTCCCGCAAGGCGCGAGGTCCTCCAGCGAGCAACCGCGCA 289
DB 61 AGGACTGTGCTTCTGAGAGGCGAGGCGCGAGGTCCTCCAGCGAGCAACCGCGCA 120

QY 290 ACAAGCCCAACAGCCGCGAGCTGAGGTGCGCGCG-----ACAACCCCGCGAGAG 343
DB 121 ACAAGCCCAACAGCCGCGAGCTGAGGTGCGCGCGCGAGAACACAGCTGAGAGAG 180

QY 344 CCGGCGCGAGGCGCGAGGCAACCTG-----AAGTTCCCGAGATCAACCTGTGAGAG 397
DB 181 CCGGCGCGAGGCGCGAGGCAACCTGAGCTTCAACTTCCCGAGATCAACCTGTGAGAG 240

QY 398 GCGCCCTGTGAGCATCAAGTGTGCGCGAGATCAAGAGGCTGTGAGACACCGCG 457
DB 241 GCGCCCTGTGAGCATCAAGTGTGCGCGAGATCAAGAGGCTGTGAGACACCGCG 300

QY 458 CCGAGCAACCGTGTGAGAGAGATGAGCTGCTCCGCGCAAGTGAAGCCCAAGATATG 517
DB 301 CCGAGCAACCGTGTGAGAGAGATGAGCTGCTCCGCGCAAGTGAAGCCCAAGATATG 360

QY 518 GCGGCAATCGCGGCTTCACTAAGTGTGCGAGTACAGACAGATCTGATGAAATCTGCG 577
DB 361 GCGGCAATCGCGGCTTCACTAAGTGTGCGAGTACAGACAGATCTGATGAAATCTGCG 420

QY 578 GCAAGAGGCGCATCGGCAACGTGTGATCGGCGCCCAACCCCGTGAACATCATCGCGCA 637
DB 421 GCAAGAGGCGCATCGGCAACGTGTGATCGGCGCCCAACCCCGTGAACATCATCGCGCA 480

QY 638 ACAATGTGATCCAGCTGAGCTGCAACCTTGAATTTCCCATCAAGCCCATGAGACGCTGC 697
DB 481 ACCTGTGATCCAGCTGAGCTGCAACCTTGAATTTCCCATCAAGCCCATGAGACGCTGC 540

QY 698 CCGTGAAGCTGAAGCCCGCGATGAGAGGCGCCCAAGTGAAGAGCTGCGCCCTGACGAGG 757
DB 541 CCGTGAAGCTGAAGCCCGCGATGAGAGGCGCCCAAGTGAAGAGCTGCGCCCTGACGAGG 600

QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGGCGCAAGATCACCA 817
DB 601 AGAAGATCAAGGCGCTGAGATCTGACCGAATGAGAGAGGCGCAAGATCACCA 660

QY 818 AGATGCGCCCGAGAACCCCTTCAACACCCCGGTTCGCTCAAGAGAGACAGCA 877
DB 661 AGATGCGCCCGAGAACCCCTTCAACACCCCGGTTCGCTCAAGAGAGACAGCA 720

QY 878 CCAAGTGTGCGCAAGCTGTGACTTCCGAGACTGAACAAGCCACCCAGAACCTTCTGCG 937
DB 721 CCAAGTGTGCGCAAGCTGTGACTTCCGAGACTGAACAAGCCACCCAGAACCTTCTGCG 780

QY 938 AGTGCAGCTGTGGCATTCCTCCACCCCGCGCTGAAGAGAGAGCGTGAACCTTC 997
DB 781 AGTGCAGCTGTGGCATTCCTCCACCCCGCGCTGAAGAGAGAGCGTGAACCTTC 840

QY 998 TGAAGTGTGCGCAAGCTTCACTTCAAGCTGCGCTTGAACAGAGACTTCCGAGTACAGG 1057
DB 841 TGAAGTGTGCGCAAGCTTCACTTCAAGCTGCGCTTGAACAGAGACTTCCGAGTACAGG 900

QY 1058 CCTTCAACATCCCGAGCATCAACAGAGACCCCGCATCCGCTACAGTACCAAGTGC 1117
DB 901 CCTTCAACATCCCGAGCATCAACAGAGACCCCGCATCCGCTACAGTACCAAGTGC 960

QY 1118 TGCCTCAAGGCTGAGAGGCAAGCCCGAGCATCTTCAAGAGAGCATGACCAAGATCTCG 1177
DB 961 TGCCTCAAGGCTGAGAGGCAAGCCCGAGCATCTTCAAGAGAGCATGACCAAGATCTCG 1020

QY 1178 AGCCCTTCCGCGCGCGCAACCCCGAGATGTATCTACAGAGGCGCCCTGTAGTGGGA 1237
DB 1021 AGCCCTTCCGCGCGCGCAAGACCCCGAGATGTATCTACAGAGGCGCCCTGTAGTGGGA 1080

QY 1238 GCGACTGTGAGATCGGCGAGCAACCGCGCAAGATGAGAGCTGTGCAAGACCTGTGCG 1297
DB 1081 GCGACTGTGAGATCGGCGAGCAACCGCGCAAGATGAGAGCTGTGCAAGACCTGTGCG 1140

QY 1298 GCTGTGGGCTTCAACACCCCGCGCAAGAGACCAAGAGAGGCGCCCTTCTGTCCAT-- 1355
DB 1141 GCTGTGGGCTTCAACACCCCGCGCAAGAGACCAAGAGAGGCGCCCTTCTGTGTATGG 1200

QY 1356 ----CGAGCTGACCCCGCAAGTGTGAGCGCTGACGCTGAGCTGCGCGAGAGAGAG 1411
DB 1201 GCTGACAGTGTGACCCCGCAAGTGTGAGCGCTGAGCGCTGACGCTGAGCTGCGCGAGAGAG 1260

QY 1412 GCTGACCGTGAACGACATCAAGAGCTGTGTGCGCAAGCTGAGCTGTGCGCGAGCAGATCT 1471
DB 1261 GCTGACCGTGAACGACATCAAGAGCTGTGTGCGCAAGCTGAGCTGTGCGCGAGCAGATCT 1320

QY 1472 ACCCGCGCATCAAGTGTGCGCGAGCTGTGCGCGCGCGCGCGCAAGCGCTTGAACG 1531
DB 1321 ACCCGCGCATCAAGTGTGCGCGAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380

QY 1532 ACATGTGCGCTTGAACCGAGAGGCGCGAGCTGTGCGCGCGCGCGCGCGCGCGCGCG 1591
DB 1381 ACATGTGCGCTTGAACCGAGAGGCGCGAGCTGTGCGCGCGCGCGCGCGCGCGCGCG 1440

QY 1592 GCGAGCGCTGACGCGGTGTACTACAGCCCGCAAGAGACTGTGTGCGCGAGATCCAG 1651
DB 1441 AGAGCGCTGTGACAGAGGTGTACTACAGCCCGCAAGAGACTGTGTGCGCGAGATCCAG 1500

QY 1652 AGCAGGCGCAAGCAAGTGTGAGCTTACAGATCTTACAGAGGCTTCAAGAACCTTGAAG 1711
DB 1501 AGCAGGCGCAAGGCGTGTGAGCTTACAGATCTTACAGAGGCTTCAAGAACCTTGAAG 1560

QY 1712 CCGGCAAGTACCGCAAGATGTGCGCAACCGCGCAACCAAGTGAAGAGCTGAGCGAGG 1771
DB 1561 CCGGCAAGTACCGCGCATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1620

QY 1772 CCGTGAAGATCGCATGTGAGAGCATGTGATCTGTGGGCAAGACCCCAAGTTCCGCG 1831
DB 1621 CCGTGAAGATGTGAGACCGAGAGCATGTGATCTGTGGGCAAGATCCCAAGTTCCGAGC 1680

QY 1298 GCTGGGCTTCAACACCCCGCAAGAACACCAAGAGAGCCCTTCTGCTCCAT-- 1355
Db 1141 GCTGGGCTTCAACACCCCGCAAGAACACCAAGAGAGCCCTTCTGCTCCAT-- 1200
QY 1356 ---CGAGCTGACCCCGCAAGTGGACCGTTCAGCCCATGAGCTGCGCGAAGAGAGA 1411
Db 1201 GCTAGAGCTGACCCCGCAAGTGGACCGTTCAGCCCATGAGCTGCGCGAAGAGAGA 1260
QY 1412 GCTGAGCCGTGACCAATCCAGAACTGTGGGCAAGCTGAACTGTGGCCAGCCAGATCT 1471
Db 1261 GCTGAGCCGTGACCAATCCAGAACTGTGGGCAAGCTGAACTGTGGCCAGCCAGATCT 1320
QY 1472 ACCCGGCAATCAAGTGGCGCCAGCTGTGCAACTGTGCGCGCCCAAGGCTTGAACCG 1531
Db 1321 AGCGCGGCAATCAAGTGGCGCCAGCTGTGCAACTGTGCGCGCCCAAGGCTTGAACCG 1380
QY 1532 ACATGCTGCGCCGCAAGCCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGC 1591
Db 1381 AGGTATCCCTGACCCGAGAGGCGGAGCTGGAGCTGGCCGAGAACCGCGAGATCTGCA 1440
QY 1592 GCGAGCCCGTGAACCGCGCTGTACTAGACCCCGCAAGAGACCTGTGGCCAGATCTGAGA 1651
Db 1441 AGGAGCCCGTGAACCGAGGCTGTACTAGACCCCGCAAGAGACCTGTGGCCAGATCTGAGA 1500
QY 1652 AGCAGGGCCAGACCACTGAGACCTTACAGATCTTACAGAGACCTTCAAGAACTTGAGA 1711
Db 1501 AGCAGGGCCAGAGGCTGAGACCTTACAGATCTTACAGAGACCTTCAAGAACTTGAGA 1560
QY 1712 CCGGCAAGTGGCCCAAGTGGCCCAAGCCCAAGCAAGCTGAAAGCAAGCTGAGCCGAGG 1771
Db 1561 CCGGCAAGTGGCCCAAGTGGCCCAAGCCCAAGCAAGCTGAAAGCAAGCTGAGCCGAGG 1620
QY 1772 CCGTCAAGAGATCCGATGAGAGACCTGTATCTGGGCAAGACCCCAAGTTCGCGC 1831
Db 1621 CCGTCAAGAGATCCGATGAGAGACCTGTATCTGGGCAAGATCCCAAGTTCAGAC 1680
QY 1832 TGCCCATCCAGAGAGAGACCTGAGAGACCTGTGTGAGACCTGAGAGGCTGAGACCTGGA 1891
Db 1681 TGCCCATCCAGAGAGAGACCTGAGAGACCTGTGTGAGAGCTGAGAGGCTGAGACCTGGA 1740
QY 1892 TCCCGGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTGCAAGCTGAGAGA 1951
Db 1741 TCCCGGAGTGGAGTTCGTGAACACCCCGCTGTGTGAAGCTGTGTGTGCAAGCTGAGAGA 1800
QY 1952 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGTGAACGCGCGCCGCAACCGGAGAGCA 2011
Db 1801 AGGAGCCCATCATCGGCGCGGAGACCTTCTAGTGAACGCGCGCCGCAACCGGAGAGCA 1860
QY 2012 AGATCGGCAAGGCGGCTTACTGTAACCGAGCCGAGGCGGAGAGATCGTGAAGCTGAGCCG 2071
Db 1861 AGCTGGGCAAGGCGGCTTACTGTAACCGAGCCGAGGCGGAGAGATCGTGAAGCTGAGCCG 1920
QY 2072 AGACCAACCAAGAGAGACCTGAGCTGAGAGCCATTCAGTGTGCTGTGAGAGAGAGGGA 2131
Db 1921 AGACCAACCAAGAGAGACCTGAGCTGAGAGCCATTCAGTGTGCTGTGAGAGAGAGGGA 1980
QY 2132 GGGAGGTGAACATCGTGAACGAGCAGGACGCTGTGGGCAATTCAGGCGCCAGCCG 2191
Db 1981 TGGAGGTGAACATCGTGAACGAGCAGGACGCTGTGGGCAATTCAGGCGCCAGCCG 2040
QY 2192 ACAAGAGCGAGAGCGAGCTGTGTGAACCAATCATTCAGAGCAAGCTGATCAAGAGAGAG 2251
Db 2041 ACAAGAGCGAGAGCGAGCTGTGTGAACCAATCATTCAGAGCAAGCTGATCAAGAGAGAG 2100
QY 2252 TGTACTGAGCTGT 2311
Db 2101 TGTACTGAGCTGT 2160
QY 2312 TGT 2371
Db 2161 TGT 2220

QY 2372 TCTACAGTATAGGAGACACCTGTAGCGGAGCGGCGCTGTAGATCATTAAGC 2431
Db 2221 TCTACAGTATAGGAGACACCTGTAGCGGAGCGGCGCTGTAGATCATTAAGC 2280
QY 2432 TTCCCGGGCTTGAACCGCGTGAATTC 2457
Db 2281 TTCCCGGGCTTGAACCGCGTGAATTC 2306
RESULT 10
AX427930 9166 bp DNA linear PAT 20-JUN-2002
LOCUS AX427930
DEFINITION Sequence 168 from Patent WO0232943.
ACCESSION AX427930
VERSION AX427930.1 GI:21538017
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
other sequences; artificial sequences.
REFERENCE
1 Huang, Y. and Nabel, G.J.
AUTHORS
Title
Modifications of hiv env, gag, and pol enhance immunogenicity for
genetic immunization
JOURNAL Patent: WO 0232943-A 168 25-APR-2002;
GOVERNMENT OF THE UNITED STATES (US)
FEATURES
source
Location/Qualifiers
1. 9166
/organism="synthetic construct"
/mol type="unassigned DNA"
/db xref="taxon:32630"
/note="plasmid pVR1012x/b containing HIV genes"
ORIGIN
Query Match 81.4%; Score 2000.8; DB 6; Length 9166;
Best Local Similarity 91.3%; Pred. No. 9,2e-164;
Matches 2185; Conservative 0; Mismatches 182; Indels 25; Gaps 5;
QY 14 TGCCGAGGCGATGAGCCAGGCGCCAGCAATCTGTATGACGCGAGCAACTTCA 73
Db 2961 TGCCGAGGCGATGAGCCAGGCGCCAGCAATCTGTATGACGCGAGCAACTTCA 3020
QY 74 AGGGCCCAAGGCGCATCATGAGCTTCAATGCGGCGCAAGAGGCGCCATGCGCCGCA 133
Db 3021 AGGGCCCAAGGCGCATCATGAGCTTCAATGCGGCGCAAGAGGCGCCATGCGCCGCA 3080
QY 134 ACTGCGGCGCCCGCGCAAGAGGCGCTGTGAGTGTGCGCAAGAGGCGCCAGATGA 193
Db 3081 ACTGCGGCGCCCGCGCAAGAGGCGCTGTGAGTGTGCGCAAGAGGCGCCAGATGA 3140
QY 194 AGACTGACCGAGCGCGAGGCGCAACTTCTTCCGAGAGACCTGTGCTTCCCGAGGCA 253
Db 3141 AGACTGACCGAGCGCGAGGCGCAACTTCTTCCGAGAGACCTGTGCTTCCCGAGGCA 3199
QY 254 AGGCGCGAGGTTCCCGAGGCGAGACCGGCGCAAGCGCCAGCGCGAGCTGCG 313
Db 3200 AGGCGCGAGGTTCCCGAGGCGAGACCGGCGCAAGCGCCAGCGCGAGGCTGCG 3259
QY 314 AGGTGCGCGG-----CGAACACCCCGCGAGAGGCGCGCGCGAGCGCGGCA--- 364
Db 3260 AGGTGCGCGGAG 3319
QY 365 ---CCCTGAATCTTCCCGAGATGACCTGTGCGAGGCGCGCGCTGTGAGATCAAGGTG 421
Db 3320 TATCTTGTAGCTTCTCTCAGATCACTTGTGCGAGGACCCCTGTGCAATAAAGATAG 3379
QY 422 GGGGCGAGTCAAGAGAGGCGCTGTGAGACCGGCGCGAGACCGGCTGTGAGAGAGA 481
Db 3380 GGGGCGAGTCAAGAGAGGCGCTTCTAGAGACCGGCGCGAGAGACCGGCTGTGAGAGAGA 3439
QY 482 TGAGCTGCGCGCGCAAGTGAAGAGCCAGAGATGATGCGCGGCTTCAATCAAG 541
Db 3440 TGAGCTGCGCGCGCGCTGTGAGAGCCAGAGATGATGCGCGGCTTCAATCAAG 3499

QY 542 TGGCCAGTACGACGAGATCTGATCGAGCTGCGCAGAAAGGCCATCGGCACCGTGC 601
DB 3500 TGGGCAAGTACGACGAGATCTGATCGAGATCTGCGGCCAGAAAGGCCATCGGCACCGTGC 3559
QY 602 TGAATGGGCCCCCCCCCGTGAATCATATGCGCGCCAGCAATGCTGACCCAGCTGGGCTGCA 661
DB 3560 TGGTGGGCCCCCCCCCGTGAATCATATGCGCGCCAGCAATGCTGACCCAGATCGCTGCA 3619
QY 662 CCTGAATCTTCCCATCAAGCCCCCATCGAGCCGTGCGCCGTGAAGCTGAAGCCCGGCAATG 721
DB 3620 CCTGAATCTTCCCATCAAGCCCCCATCGAGCCGTGCGCCGTGAAGCTGAAGCCCGGCAATG 3679
QY 722 AGCGCCCCAGGTGAAGCGAGTGGCCCTTGAACGAGAGAAATCAAGGCCCTGACGCCCA 781
DB 3680 AGCGCCCCAGGTGAAGCGAGTGGCCCTTGAACGAGAGAAATCAAGGCCCTGAGTGAAGA 3739
QY 782 TCTTGGAGAGATGAGAGAGAGAGGCGAGATCAACCAATTCGCCCCCGAGAAACCCCTTACA 841
DB 3740 TCTTGAACCGAGATGAGAGAGAGAGGCGAGATCAACCAATTCGCCCCCGAGAAACCCCTTACA 3799
QY 842 ACAACCCCGCTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 3800 ACAACCCCGCTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3859
QY 902 TCCGAGAGCTGAACAGCGCACCCAGAGATCTTCTGGAGAGTGCAGCTGGGCAATCCCCAC 961
DB 3860 TCCGAGAGCTGAACAGCGCACCCAGAGATCTTCTGGAGAGTGCAGCTGGGCAATCCCCAC 3919
QY 962 CGCGCGGCTTGAAG 1021
DB 3920 CGCGCGGCTTGAAG 3979
QY 1022 GCGTCCCGCTGAGACGAGATCTTCCGAGAGTACAGCGCTTCAACATCCAGATCAACA 1081
DB 3980 GCGTCCCGCTGAGACGAGATCTTCCGAGAGTACAGCGCTTCAACATCCAGATCAACA 4039
QY 1082 ACAGAGACCCCGCGCATCGCTACAGATCAAGTCAAGTCTGCGCCAGAGCTGGAAGGCAAGC 1141
DB 4040 ACAGAGACCCCGCGCATCGCTACAGATCAAGTCAAGTCTGCGCCAGAGCTGGAAGGCAAGC 4099
QY 1142 CGAGCATCTTCCAGAGAGAGATCAACAGATCTTGAAGCCCTTCCGCGCCGCAACCCCG 1201
DB 4100 CGGCAATCTTCCAGAGAGAGATCAACAGATCTTGAAGCCCTTCCGCGCGCAACCCCG 4159
QY 1202 AGATGCTGATCTTACCA-----GCCCCCTGTGATGCTGGGCAAGCACTGAGATCGGC 1255
DB 4160 ACATGCTGATCTTACCAAGTCAATGACACACTGTAAGTGGAGCAAGCTTGAAGATCGGC 4219
QY 1256 AGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGCTTCAACCAACC 1315
DB 4220 AGCACCGCGCAAGATCGAGAGCTGCGCAAGCACTGCTGCGCTGGGCTTCAACCAACC 4279
QY 1316 CGGACAAAGAGACCAAG 1369
DB 4280 CGGACAAAGAGACCAAG 4339
QY 1370 ACAAGTGAACGCTGACAGCCATCGAGAGCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1429
DB 4340 ACAAGTGAACGCTGACAGCCATCGTGTCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4399
QY 1430 TCCAGAGAGTGGTGGGCAAGCTGAATCTGGGCGAGCCAGCATCTTACCCCGCATCAAGTGC 1489
DB 4400 TCCAGAGAGTGGTGGGCAAGCTGAATCTGGGCGAGCCAGCATCTTACCCCGCATCAAGTGC 4459
QY 1490 GCCAGCTGTGCAAGCTGCTGCGCGCGCGCAAGGCCCTTGAACCGACATCTGTGCCCTGACCG 1549
DB 4460 GCCAGCTGTGCAAGCTGCTGCGCGCGCGCAAGGCCCTTGAACCGACATCTGTGCCCTGACCG 4519
QY 1550 AGAGAGCGAGCTGAGAGCTGCGCGCGCGCAAGGCCCTTGAACCGAGAGAGAGAGAGAGAGAGAG 1609
DB 4520 AGAGAGCGAGCTGAGAGCTGCGCGCGCGCAAGGCCCTTGAAGAGAGAGAGAGAGAGAGAGAGAG 4579
QY 1610 TGTACTTACGACCCAGCAAG 1669

DB 4580 TGTACTTACGACCCAGCAAG 4639
QY 1670 GAACTTACCAATCTTACAG 1729
DB 4640 GAACTTACCAATCTTACAG 4699
QY 1720 TGGCGACCGGCCCAACCAAG 1789
DB 4700 TGAAGAGGATGATGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4759
QY 1790 TGAAGAGATGATGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1849
DB 4760 CGAGAGAGATGATGATCTGGGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4819
QY 1850 CTTGGAG 1909
DB 4820 CTTGGAG 4879
QY 1910 TGAACACCCCGCTGTGAG 1969
DB 4880 TGAACACCCCGCTGTGAG 4939
QY 1970 CGAG 2029
DB 4940 CGAG 4999
QY 2030 AGGTGACCGAG 2089
DB 5000 AGGTGACCGAG 5059
QY 2090 CGAG 2149
DB 5060 CGAG 5119
QY 2150 CGAG 2209
DB 5120 CGAG 5179
QY 2210 TGGTGAACCAATCTTACAG 2269
DB 5180 TGGTGAACCAATCTTACAG 5239
QY 2270 CGGCGCAAG 2329
DB 5240 CGGCGCAAG 5299
QY 2330 GCAAGTGTCTTCTGAG 2381
DB 5300 GCAAGTGTCTTCTGAG 5351

RESULT 11
BD263706 2312 bp DNA linear PAT 17-JUL-2003
LOCUS BD263706
DEFINITION Improved expression of HIV polypeptides and production of
virus-like particles.
ACCESSION BD263706.1 GI:33073474
VERSION JP 200253124-A/73.
KEYWORDS synthetic construct
SOURCE
ORGANISM
other sequences; artificial sequences.
REFERENCE
1 (bases 1 to 2312)
Barnett,S., Megede,J.Z., Sriwastava,I., Lian,Y., Hartog,K., Liu,H.,
Greer,C., Selby,M. and Walker,C.
Improved expression of HIV polypeptides and production of
virus-like particles
JOURNAL
CHIRON CORP
OS Artificial Sequence
PN JP 200253124-A/73
PD 08-OCT-2002

PF 30-DEC-1999 JP 2000591193
 PR 31-DEC-1998 US 60/114495, 01-DEC-1999 US 60/168471, PI
 SUSAN BARRETT, JAN ZUR MEGEDE, INDRESH SRIVASTAVA, YING LIAN, PI
 KARIN HARTOG,
 PI HONG LIU, CATHERINE GREER, MARK SELBY, CHRISTOPHER WALKER, PC
 C12N15/09, A61K31/711, A61K38/00, A61K48/00, A61P31/18, A61P37/02, PC
 C12N5/10,
 PC C12N7/00, C12P21/02, C12N15/00, C12N5/00, A61K37/02 CC
 Description of Artificial Sequence: FS(-). protomod. Rtopc (+) FH
 Key
 FT source 1. .2312
 Location/Qualifiers
 FEATURES
 source 1. .2312
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"

ORIGIN

Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
 Best Local Similarity 92.9%; Pred. No. 2,4e-163;
 Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

170 GGGGAGAGGGGCGCAGATGAAGACTGACCGAGGGCCAGGCGCACTTCCCGG 229
 1 GCGGCGCGAGAGCACAATAAAGATTGCACTGAGAGACAGGCTAATTCTTCGCG 60
 230 AGGACCTGAGCTTCCCGCAGAGCGCGAGGCGGAGTTCCCGAGCGAGAAACCGCGCA 289
 61 AGGACCTGAGCTTCCCGCAGAGCGCGAGGCGGAGTTCCCGAGCGAGAAACCGCGCA 120
 290 AAGACCCCAACGAGCGCGAGCTGACAGTGGCGGCG-----ACAACCCCGCAGGAGG 343
 121 AAGACCCCAACGAGCGCGAGCTGACAGTGGCGGCGCGAGAAACAACCTGAGCGAGG 180
 344 CGGAGCGAGGCGCGAGGCGCGAGCTG-----AACTTCCCGCAGATCACTGTGCGAG 397
 181 CGGAGCGCGAGGCGCGAGGCGCGAGCTTCACTTCCCGCAGATCACTGTGCGAG 240
 398 GCGCCCTGTGAGCATCAAGTGGCGCGAGATCAAGAGGCGCTGTGAGCAACCGCGG 457
 241 GCGCCCTGTGAGCATCAAGTGGCGCGAGCTGAGAGGCGCTGTGAGCAACCGCGG 300
 458 CGGAGCATCCGTGTGAGAGAGATGAGCTGCGCGCAAGTGAAGCCCAAGATGATG 517
 301 CGGAGCATCCGTGTGAGAGAGATGAACTGCGCGCAAGTGAAGCCCAAGATGATG 360
 518 GCGGATCGGCGGCTTCAATCAAGTGGCGCGAGATCAAGAGGCGCTGTGAGTGGG 577
 361 GCGGATCGGCGGCTTCAATCAAGTGGCGCGAGATCAAGAGGCGCTGTGAGTGGG 420
 578 GCAAGAGGCGCATCGGACCGTGTGATGGCGCGCAACCCCGTGAACATCATCGGCGCA 637
 421 GCGCAAGGCGCATCGGACCGTGTGATGGCGCGCAACCCCGTGAACATCATCGGCGCA 480
 638 ACAATGTGAACCGAGCTGGGCTGCAACCTGAATTCCCATCAGCCCATGAGACCGTGC 697
 481 ACTGTGTGAACCGAGCTGGGCTGCAACCTGAATTCCCATCAGCCCATGAGACCGTGC 540
 698 CCGTGAAGCTGAAGCGCGGAGTGAAGCGCGCAAGGTGAAGAGCGTGGCCCTTACCGAGG 757
 541 CCGTGAAGCTGAAGCGCGGAGTGAAGCGCGCAAGGTGAAGAGCGTGGCCCTTACCGAGG 600
 758 AGAAGATCAAGGCGCTGACCGCATCTGACGAGAGATGAGAGAGAGGCGCAAGTACCA 817
 601 AGAAGATCAAGGCGCTGACCGCATCTGACGAGAGATGAGAGAGAGGCGCAAGTACCA 660
 818 AGATGGGCGCGAGAAACCTTCAACAACAACCCCGTGTTCGCATCAAGAGAGAGACGA 877
 661 AGATGGGCGCGAGAAACCTTCAACAACAACCCCGTGTTCGCATCAAGAGAGAGACGA 720
 878 CCAAGTGGCGCGAGCTGTGAGATTCGCGAGAGTGAACAGCGCACCCGAGACTTCTGGG 937

721 CCAAGTGGCGAGCTGTGAGACTTCGCGAGCTGAACAAGGCGCACCCAGACTTCTGGG 780
 938 AGGTCACTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGGTGACCGTGC 997
 781 AGGTCACTGGGCAATCCCCCAACCCCGCGGCTTGAAGAGAGAGAGCGGTGACCGTGC 840
 998 TGAAGTGGGCGAGCGCTACTTCACTGAGCGGCGCTGAGACGAGGACTTCCGCAAGTACAGG 1057
 841 TGAAGTGGGCGAGCGCTACTTCACTGAGCGGCGCTGAGACGAGGACTTCCGCAAGTACAGG 900
 1058 CTTTACCATCCCGCAGATCAACAAGAGACCCCGGCAATCCGCTACCAAGTGC 1117
 901 CTTTACCATCCCGCAGATCAACAAGAGACCCCGGCAATCCGCTACCAAGTGC 960
 1118 TCCCGCAGGCGCTGGAAGGCGAGCCCGCAGATTTTCAAGAGCGATGACCAAGTCTCG 1177
 961 TCCCGCAGGCGCTGGAAGGCGAGCCCGCAGATTTTCAAGAGCGATGACCAAGTCTCG 1020
 1178 AGCCCTTCGCGCGCGCGCAACCCCGAGATGTGATCAACA-----GGCCCGCTGTAG 1231
 1021 AGCCCTTCGCGCGCGCGCAACCCCGAGATGTGATCAACAAGTGTGAGACCTGTAG 1080
 1232 TGGGAGCGACCTTGAAGATCGGCGAGCAACCGCGCAAGATCGAGAGCTGCGCAAGCAC 1291
 1081 TGGGAGCGACCTTGAAGATCGGCGAGCAACCGCGCAAGATCGAGAGCTGCGCGCAAGCAC 1140
 1292 TCTGCGCTGGGCTTCAACCAACCCCGCAGAGAGCAACAAGAGACCCCTTCTGCG 1351
 1141 TCTGCGCTGGGCTTCAACCAACCCCGCAGAGAGCAACAAGAGACCCCTTCTGCG 1200
 1352 CCAT-----CGAGTGCACCCCGCAGAGTGAACCGTGAAGCCCATCGAGCTGCCGAG 1405
 1201 GATGAGGCTTCAAGAGTGCACCCCGCAGAGTGAACCGTGAAGCCCATCGAGCTGCCGAG 1260
 1406 AGAGAGCTGAGACCTGTAACGATCAATCAAGACTGTGGGCAAGTGAACCTGGGCGAGCC 1465
 1261 AGAGAGCTGAGACCTGTAACGATCAATCAAGACTGTGGGCAAGTGAACCTGGGCGAGCC 1320
 1466 AGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCC 1525
 1321 AGATCTACCCCGCATCAAGTGGCGCAGCTGTGCAAGCTGTGCGCGCGCGCAAGGCC 1380
 1526 TGAACGCAATCGTCCCGTCAACGAGAGGCGGAGCTGAGAGCTGCGCGAGAACCCGCGAG 1585
 1381 TGAACGCAATCGTCCCGTCAACGAGAGGCGGAGCTGAGAGCTGCGCGAGAACCCGCGAG 1440
 1586 TCTGTGCGAGCGCGGCGAGCTGAGTCAAGACCCCGCAAGAGCTGTGTGCGAGAG 1645
 1441 TCTGTGCGAGCGCGGCGAGCTGAGTCAAGACCCCGCAAGAGCTGTGTGCGAGAG 1500
 1646 TCCAGAGAGGCGCGCAACGAGTGAACCTTACCAAGTCTTCAAGAAC 1705
 1501 TCCAGAGAGGCGCGCAACGAGTGAACCTTACCAAGTCTTCAAGAAC 1560
 1706 TGAAGACCGGCAAGTACGCGCAAGTGCACCGCGCGCAACCAAGCTGGAAGCACTGA 1765
 1561 TGAAGACCGGCAAGTACGCGCGAGTGCAGCGCGCGCGCAACCAAGCTGGAAGCACTGA 1620
 1766 CCGAGGCGGTGAAGAGATCGCATGGAAGAGATGTGATCTGGGCGAGAACCCCGAGT 1825
 1621 CCGAGGCGGTGAAGAGATCGCATGGAAGAGATGTGATCTGGGCGAGAACCCCGAGT 1680
 1826 TCCGCTGCGCATTCAGAGAGAGACTTGAAGAGACTTGAAGACCACTTGAAGAGCGCA 1885
 1681 TCAAGTGCCTTCAAGAGAGAGACTTGAAGAGACTTGAAGAGACTTGAAGAGCGCA 1740
 1886 CTTGATCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTAAGCTGTGTACCAAC 1945
 1741 CTTGATCCCGAGTGGAGGTTCTGTAACAACCCCGCTGTGTAAGCTGTGTACCAAC 1800
 1946 TGAAGAGAGCGCATCATCGGCGCGAGACTTCTTCAAGTGAACGCGCGCGCAACCGCG 2005
 1801 TGAAGAGAGCGCATCATCGGCGCGAGACTTCTTCAAGTGAACGCGCGCGCAACCGCG 1860

QY 2006 AGACCAAGATCGGCAAGGCGGCTGATGACCGGCGCGGAGAGATGCGAGCC 2065
 DB 1861 AGACCAAGATCGGCAAGGCGGCTGATGACCGGCGCGGAGAGATGCGAGCA 1920
 QY 2066 TGAACGAGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2125
 DB 1921 TCGCGGACGACCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
 QY 2126 GCGGCAAG 2185
 DB 1981 GCGGCAAG 2040
 QY 2186 AGCGGCAAG 2245
 DB 2041 AGCGGCAAG 2100
 QY 2246 AGAGAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2305
 DB 2101 AGAGAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
 QY 2306 AGAGAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2365
 DB 2161 AGAGAGATGATCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
 QY 2366 TCGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2425
 DB 2221 TCGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2280
 QY 2426 AAAAGCTTCCCGGAGCTAGACCGGATGATTC 2457
 DB 2281 AAAAGCTTCCCGGAGCTAGACCGGATGATTC 2312

RESULT 12
 LOCUS CO870576 2312 bp DNA linear PAT 13-SBP-2004
 DEFINITION Sequence 84 from Patent EP1433851.
 ACCESSION CO870576
 VERSION CO870576.1 GI:52000092

KEYWORDS
 SOURCE
 ORGANISM
 other sequences; artificial sequences.
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 CHIRON CORPORATION (US)

FEATURES
 source
 1..2312
 /organism="synthetic construct"
 /mol_type="unassigned DNA"
 /db_xref="taxon:32630"
 /note="Description of Artificial Sequence:
 FS(-).promod.RToc(+)"

ORIGIN
 Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
 Best Local Similarity 92.9%; Pred. No. 2.4e-163;
 Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;

QY 170 GCGGCAAG 229
 DB 1 GCGGCGCGCAAG 60
 QY 230 AGGACCTGAGCTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 289
 DB 61 AGGACCTGAGCTTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
 QY 290 AAGAGCCCAAG 343

DB 121 AAGAGCCCAAG 180
 QY 344 CCGGCGCGAG 397
 DB 181 CCGGCGCGAG 240
 QY 398 GCGGCGCGAG 457
 DB 241 GCGGCGCGAG 300
 QY 458 CCGAG 517
 DB 301 CCGAG 360
 QY 518 GCGGAGATGAG 577
 DB 361 GCGGAGATGAG 420
 QY 578 GCGAG 637
 DB 421 GCGAG 480
 QY 638 AGATGAG 697
 DB 481 AGATGAG 540
 QY 698 CCGGAG 757
 DB 541 CCGGAG 600
 QY 758 AGAGAGATGAG 817
 DB 601 AGAGAGATGAG 660
 QY 818 AGATGAG 877
 DB 661 AGATGAG 720
 QY 878 CCGAG 937
 DB 721 CCGAG 780
 QY 938 AGGAG 997
 DB 781 AGGAG 840
 QY 998 TGAAG 1057
 DB 841 TGAAG 900
 QY 1058 CTTTCAAG 1117
 DB 901 CTTTCAAG 960
 QY 1118 TGGGCAAG 1177
 DB 961 TGGGCAAG 1020
 QY 1178 AGGCTTCCCGGAG 1231
 DB 1021 AGGCTTCCCGGAG 1080
 QY 1232 TGGGCAAG 1291
 DB 1081 TGGGCAAG 1140
 QY 1292 TGGGCAAG 1351
 DB 1141 TGGGCAAG 1200
 QY 1352 CCAAT-----GAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1405

Db 1201 GGATGGGCTACGAGCTGCAACCCCGACAAGTGGACCGTGCAGCCCATCTGCTCCCGAGA 1260
QY 1406 AGGAGAGCTGGAACCTGTGAACGATTCAGAGAGCTGTGTGGCAAGCTGAACTGTGGCCAGCC 1465
Db 1261 AGGACAGCTGGAACCGTGAACGATTCAGAGAGCTGTGTGGCAAGCTGAACTGTGGCCAGCC 1320
QY 1466 AGATCTACCCCGGCACTCAAGAGTGTGGCCAGAGCTGTGCAAGCTGTGTGGCCGCGCAAGGCC 1525
Db 1321 AGATCTACCCCGGCACTCAAGAGTGTGGCCAGAGCTGTGCAAGCTGTGTGGCCGCGCAAGGCC 1380
QY 1526 TGAACGACATCGTGCCTGCTGACCGAGAGAGCGAGCTGTGAGCTGTGGCCGAGAACCGCGAGA 1585
Db 1381 TGAACGAGGTGATTCCTGCTGACCGAGAGAGCGAGCTGTGAGCTGTGGCCGAGAACCGCGAGA 1440
QY 1586 TCCTGCGAGAGCCCGTGCACGCGGTGTACTACGACCCCAAGAGAGCTGTGTGGCCGAGA 1645
Db 1441 TCCTGAGAGAGCCCGTGCACGAGGTGTACTACGACCCCAAGAGAGCTGTGTGGCCGAGA 1500
QY 1646 TCCAGAGAGAGGCGCCAGACGAGTGAACCTTACCAATCTACAGAGAGCCCTTCAAGAAC 1705
Db 1501 TCCAGAGAGAGGCGCCAGAGCTGTGACCTTACCAATCTACAGAGAGCCCTTCAAGAAC 1560
QY 1706 TGAAGACCGGAGATACGCGCAAGATGCGACCGCCCAACCAAGAGCTGAAGAGCTGA 1765
Db 1561 TGAAGACCGGAGATACGCGCGCATGCGCGGCCCAACCAAGAGCTGAAGAGCTGA 1620
QY 1766 CCGAGGCGCTGTGAGAGATGCGCATGAGAGCATGCTGTGTGTGTGGGCAAGACCCCAAGT 1825
Db 1621 CCGAGGCGCTGTGAGAGATGCGCATGAGAGCATGCTGTGTGTGTGGGCAAGATCCCAAGT 1680
QY 1826 TCCGCTGCTGCTTCCAGAGAGAGACTGTGGAGACTGTGTGACCCACTGTGGCAGGCCA 1885
Db 1681 TCAAGGTGCTTCCAGAGAGAGACTGTGGAGACTGTGTGATGTGATGTGATCTGGCAGGCCA 1740
QY 1886 CCGTGAATCCCGAGTGGAGATGTGTGAACACCCCGCTGTGAAGCTGTGGTACAGC 1945
Db 1741 CCGTGAATCCCGAGTGGAGATGTGTGAACACCCCGCTGTGAAGCTGTGGTACAGC 1800
QY 1946 TGAAGAGAGAGCCCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCGCCAGACCGCG 2005
Db 1801 TGAAGAGAGAGCCCATCTGCGCGCGAGACCTTCTACGTGAGCGCGCGCCAGACCGCG 1860
QY 2006 AGACCAAGTGTGGCAAGGCGCGCTACGTGAACCGACCGGCGCGCGCAAGAGTGTGAGCC 2065
Db 1861 AGACCAAGTGTGGCAAGGCGCGCTACGTGAACCGACCGGCGCGCGCAAGAGTGTGAGCA 1920
QY 2066 TGAACGAGACCAACCAAGAGAGCGAGTGTGAGCGCATTCACGCTGTGGCCCTGCAAGACA 2125
Db 1921 TGAACGAGACCAACCAAGAGAGCGAGTGTGAGCGCATTCACGCTGTGGCCCTGCAAGACA 1980
QY 2126 GCGGCAAGAGAGTGAACATCTGTGACCGAGCGAGCATTCGCTGTGGCATCTACAGGCC 2185
Db 1981 GCGGCAAGAGAGTGAACATCTGTGACCGAGCGAGCATTCGCTGTGGCATCTACAGGCC 2040
QY 2186 AGTCCGACAAAGCGAGCGAGCGAGCTGTGAACCAATCTACGAGCGCTGTATCAAGAG 2245
Db 2041 AGTCCGACAAAGCGAGCGAGCGAGCTGTGAACCAATCTACGAGCGCTGTATCAAGAG 2100
QY 2246 AGAAGGTATACCTGAGCTGTGTGCGCGCCCAAGAGGCAATCGCGCGCAACGAGCAGATG 2305
Db 2101 AGAAGGTATACCTGAGCTGTGTGCGCGCCCAAGAGGCAATCGCGCGCAACGAGCAGATG 2160
QY 2306 ACAAGCTGTGAGCAAGGCGCATTCGCAAGGTGTCTGTCTGTGACCGGCATCTGATGGCGCA 2365
Db 2161 ACAAGCTGTGAGCGCGCGCATTCGCAAGGTGTCTGTCTGTGACCGGCATCTGATGGCGCA 2220
QY 2366 TCGTGTATCTACGATACATGAGAGAGCTGTGAGCTGTGGCAGCGCGCGCTTGAATGATTT 2425
Db 2221 TCGTGTATCTACGATACATGAGAGAGCTGTGAGCTGTGGCAGCGCGCGCTTGAATGATTT 2280
QY 2426 AAAAGCTTCCCGGCGCTAGACCGGTAATTC 2457
Db 2281 AAAAGCTTCCCGGCGCTAGACCGGTAATTC 2312

RESULT 13
AR373389 2312 bp DNA linear PAT 18-DEC-2003
LOCUS AR373389
DEFINITION Sequence 84 from patent US 6602705.
ACCESSION AR373389
VERSION AR373389.1 GI:40075492
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2312)
AUTHORS Barnett,S.W., Megede,J., Greer,C. and Selby,M.
TITLE Expression of HIV polypeptides and production of virus-like particles
JOURNAL Patent: US 6602705-A 84 05-AUG-2003;
Chiron Corporation; Emeryville, CA
FEATURES
source
location/Qualifiers
1..2312
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Query Match 81.3%; Score 1998.4; DB 6; Length 2312;
Best Local Similarity 92.9%; Pred. No. 2.4e-163;
Matches 2147; Conservative 0; Mismatches 141; Indels 24; Gaps 4;
QY 170 GCGGCAAGAGGCGCCAGATGAAGAGACTGACCGAGCGCCAGCCACTTCTTCCGCG 229
Db 1 GCGGCGCGGAGAGACCAAAATGAATGTGACTGAGAGAGAGGTATTTCTTCCGCG 60
QY 230 AGAAGCTGTGCTTCTCCCAAGGCGCGGAGTTCCTCCAGCGAGCAAAACCGCGCA 289
Db 61 AGAAGCTGTGCTTCTCCAGAGGCGCGGAGTTCAGAGCGAGCGCGCGCA 120
QY 290 AAGGCGCCAGCGCGCGAGCTGCAAGTGGCGGGG-----ACAACCCCGCAGGAG 343
Db 121 AAGGCGCCAGCGCGCGAGCTGCAAGTGTGGCGCGCGAGAACACAGCTGAGCGAG 180
QY 344 CCGGCGCGAGGCGCCAGGCGCACCTG-----AATTCCCGAGATCACTGTGCGAGC 397
Db 181 CCGGCGCGAGGCGCCAGGCGCACCTGAGCTTCACTTCCCGAGATCACTGTGCGAGC 240
QY 398 GCCCCTGTGTGAGCACTAAGTGTGGCGCGCAGATCAAGAGGCGCTGTGACACCGGCG 457
Db 241 GCCCCTGTGTGACATCAAGATCGGCGCGCAGCTCAAGAGGCGCTGTGACACCGGCG 300
QY 458 CCGAGCAACCGTGTGTGAGAGAGTGAAGCTGTGCGCGCGCAAGTGAAGCCCAAGATGATCG 517
Db 301 CCGAGCAACCGTGTGTGAGAGAGTGAAGCTGTGCGCGCGCAAGTGAAGCCCAAGATGATCG 360
QY 518 GCGGCAATCGGCGGCTTCACTCAAGGTGTGCGCGCAGTACAGACAGATCTGTGAGATCTGCG 577
Db 361 GCGGCAATCGGCGGCTTCACTCAAGGTGTGCGCGCAGTACAGACAGATCTGTGAGATCTGCG 420
QY 578 GCAAGAGGCGCATTCGCGCAGTGTGAGCGCGCCACCCCGGTGAACATCTGAGCGCGCA 637
Db 421 GCAAGAGGCGCATTCGCGCAGTGTGAGCGCGCCACCCCGGTGAACATCTGAGCGCGCA 480
QY 638 ACATGTGACCCAGCTGTGAGTGAACCTTCCCATCAAGCCCATCGAGACCGTGC 697
Db 481 ACATGTGACCCAGATTCGAGTGAACCTTCCCATCAAGCCCATCGAGACCGTGC 540
QY 698 CCGTGAAGCTGAAGCCCGGCAATGAGCGCGCCCAAGGTGAAGAGTGTGCGCGAGCGAG 757
Db 541 CCGTGAAGCTGAAGCCCGGCAATGAGCGCGCCCAAGGTGAAGAGTGTGCGCGAGCGAG 600
QY 758 AGAAGATCAAGGCGCTGACCGCATCTGCGAGAGATGAGAGAGAGGCGCAAGATCAACA 817
Db 601 AGAAGATCAAGGCGCTGAGTGAATCTGCAACGAGATGAGAGAGGCGCAAGATCAACA 660
QY 818 AGATCGGCGCCGAGAACCCCTACACACCCCGTGTTCGCGCATCAAGAGAGAGACAGCA 877

D	3021	AGGGCCCAAGCGACCATCAATGATCTTCAACTGCGGCAAGAGGGCCACTCGGCCGA	3080
Q	134	ACTGCGCGCGCCCCCGCAAGAGGGCTGTGAAATGTGCGGCAAGAGGCCACCAATGA	193
D	3081	ACTCGCGGCGCCCCCGCAAGAGGGCTGTGAAATGTGCGGCAAGAGGGCCACTCGGTGA	3140
Q	194	AGGACTGCAACCGACCGCCAGGCTCACTTCTTCGCGAGAGACTGTGGCTTCTCCCAAGGCA	253
D	3141	AGGACTGCAACCGACCGCCAGGCTCACTTCTTCGCGAGAGACTGTGGCTTCTCCCAAGGCA	3194
Q	254	AGGCGCGGAGTTTCCCCGACGAGCAAGAACCGCGCAACAGGCCCAACGCGCGAGCTGC	313
D	3195	AGGCGGAGGAAATTTTCTTCAAGACGACGACCAAGCCCAAGCCCAACGAAGAGGCTTC	3254
Q	314	AGGTGCGCGG-----CGACAAACCCCGCGACGAGCGCGCGCGCGCAAGGCA---	364
D	3255	AGGTTTGGGGAAGAAGACAACAATCTCCCTTCAGAAAGCAGAGCGGATGACAAAGAACTG	3314
Q	365	---CCCTGAATTTCCCGCAAGTCAACCTGTGTGGCAGCGCCCTGTGTAGCATCAAGTGG	421
D	3315	TATCTTTAGCTTCCCTCAGATCACTTTTGGCAGCGACCCCTGTCACATTAAGATAG	3374
Q	422	GGGCGCAGATCAAGAGAGGCGCTGTGCGACAACGCGCGCGACAGACACCGTGTGAGGGA	481
D	3375	GGGCGCAGCTAAGAGAGGCGCTTCTAGACACCGGCGCGACAGACACCGTGTGAGGGA	3434
Q	482	TGAGCTGTCCCGGCAAGTGAAGGCCCAAGATGATCGCGCGCATCGCGGCTTCAAG	541
D	3435	TGAACCTGTCCCGCGCGCTGGAAGGCCCAAGATGATCGCGCGCATCGCGGCTTCAAG	3494
Q	542	TGCGCGAGTACGACCAAGATCTGTATCGAGATCTGTGCGGCAAGAGGCCATCGGCAACCTGTC	601
D	3495	TGGGCGAGTACGACCAAGATCTGTATCGAGATCTGTGCGGCGCAAGAGGCCATCGGCAACCTGTC	3554
Q	602	TGATTCGGGCGGCAACCCCGGTGAACATCATTCGCGCGCGGAATGCTCTGACCCAGCTGGGCTGCA	661
D	3555	TGTGTGGGCGGCAACCCCGGTGAACATCATTCGCGCGCGGAACCTGTGACCCAGATCGGCTGCA	3614
Q	662	CCCTGAATCTTCCCATCAGCCCCCATTCAGACCGGTGCGCGTGAAGCTGAAGCCCGGCAATG	721
D	3615	CCCTGAATCTTCCCATCAGCCCCCATTCAGACCGGTGCGCGTGAAGCTGAAGCCCGGCAATG	3674
Q	722	ACGGCGCCCAAGTGAAGCATGTGGCCCTTGACCGAGAGAAAGATCAAGGCCCTTGACCGGCA	781
D	3675	ACGGCGCCCAAGTGAAGCATGTGGCCCTTGACCGAGAGAAAGATCAAGGCCCTTGAGTGAAGA	3734
Q	782	TCTGCGAGAGATGAGAGAGAGGGCAAGATCAACAAATGTGGCCCGGAGAACCCCTTACA	841
D	3735	TCTGCAACCGAATGAGAGAGAGGGCAAGATCAACAAATGTGGCCCGGAGAACCCCTTACA	3794
Q	842	ACACCCCGGTGTGCGCATCAAGAAAGAGGACAGACCAAGTGGCGGCAAGCTGTGTGACT	901
D	3795	ACACCCCGGTGTGCGCATCAAGAAAGAGGACAGACCAAGTGGCGGCAAGCTGTGTGACT	3855
Q	902	TCCGGAAGTGAACAAGGCAACCGACATTCGTGGAGTGCAGCTTGGGCAATCCCAAC	961
D	3855	TCCGGAAGTGAACAAGGCAACCGACATTCGTGGAGTGCAGCTTGGGCAATCCCAAC	3914
Q	962	CCGCGGCGCTGAAGAAAGAAAGCGTGAACCGTGTGAGAGTGGGCAACGCTTACTTCA	1021
D	3915	CCGCGGCGCTGAAGAAAGAAAGCGTGAACCGTGTGAGAGTGGGCAACGCTTACTTCA	3974
Q	1022	GCGTGGCGCTGGAAGAGACTTTCGCAAGTACACCGGCTTCAACATCCCAAGATCAACA	1081
D	3975	GCGTGGCGCTGGAAGAGACTTTCGCAAGTACACCGGCTTCAACATCCCAAGATCAACA	4034
Q	1082	ACGAGACCCCGGCAATCGGCTTACAGTACAAAGTGTGCTGCCCAAGGCGTGAAGGCGCACCC	1141
D	4035	ACGAGACCCCGGCAATCGGCTTACAGTACAAAGTGTGCTGCCCAAGGCGTGAAGGCGCACCC	4094
Q	1142	CCAGGATCTTCAAGACGACATGACCAAGATCTTGAAGCCTTTCGCGCGCGCAACCCCG	1201
D	4095	CCGCGATCTTCAAGACGACATGACCAAGATCTTGAAGCCTTTCGCGCGCGCAACCCCG	4154

QY	1202	AGATCGGATCTTACCA-----GGCCCCCTGTACCGGGCAGGACCTTGAATCGGAC	1255
Db	4155	ACATGTGATCTTACCAATGACCACTGTACGTGGCAGCGACTTGGATCGGCC	4214
QY	1256	AGCACCGCGCAAGATCGAGAGCTGCGCAAGCACCTGTGCTGGGCTTCAACACC	1315
Db	4215	AGCACCGCACCAAGATGAGAGAGCTGGCCAGCACCTGCTGGGGCTTCAACACC	4274
QY	1316	CCGACAAAGAACACAAAGAGCCCCCTTCTCCGCCAT-----CGACTGCACCCG	1369
Db	4275	CCGACAAAGAGCACAAAGAGAGCCCCCTTCTCTGGAGATGGGCTACGACTGCACCCG	4334
QY	1370	ACAAGTGAACCGTGCACAGCCCATCGAGCTGCACCGAAGAGAGAGCTGCACCGTGAACGACA	1429
Db	4335	ACAAGTGAACCGTGCACAGCCCATCGATGTGTGCCCCGAGAGAGACAAGTGCACCGTGAACGACA	4394
QY	1430	TCCAGAGCTGTGTGGGCAAGCTGAATTGGGCAAGCCAGACTCTACCCCGGACATCAAGTGC	1489
Db	4395	TCCAGAGCTGTGTGGGCAAGCTGAATTGGGCAAGCCAGACTCTACCGCGGACATCAAGGTGC	4454
QY	1490	GCCAGCTGTCAAGCTGTCCGGGGCGCCAAAGGCCAGACCGACATCGTGCCTTCAGCG	1549
Db	4455	GCCAGCTGTCAAGCTGTCCGGGGCGCCAAAGGCCCTTGAACGAGTGTGTCCCTTGACCG	4514
QY	1550	AGGAGGCCGAGCTGGAAGCTGGCCGAGAACCCGAGATCTTACGCGAGACCCGTGCACGGCG	1609
Db	4515	AGGAGGCCGAGCTGGAAGCTGGCCGAGAACCCGAGATCTTGAAGAGACCCGTGCACGGCG	4574
QY	1610	TGTACTACGACCCCGACAAAGACCTTGGTGGCCGAGATTCAGAAAGCAGGCGCACCAAGT	1669
Db	4575	TGTACTACGACCCCGACAAAGACCTTGAATGCCCGAGATTCAGAAAGCAGGCGCCAGT	4634
QY	1670	GGAACCTACCAAGATCTTACAGAGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCAAGA	1729
Db	4635	GGAACCTACCAAGATCTTACAGAGAGCCCTTCAAGAACTGAAGACCGGCAAGTACGCCCGCA	4694
QY	1730	TGCGCACCGCCCAACCAAGACGTGAAGACCTGACCGAGGCCGTGCAGAGAGTTCGCA	1789
Db	4695	TGAAGGGCGGCCAACACCAAGACGTGAAGACGTGACCGAGGCCGTGCAGAGAGTTCGCCA	4754
QY	1790	TGAGAGACATCTGTGATCTTGGGGGCAAGACCCCAAGTTCCGCTGGCCCATCAAGAGAGA	1849
Db	4755	CCGAGACATCTGTATCTTGGGGCAAGACCCCAAGTTCAAGCTGCCCATCAAGAGAGA	4814
QY	1850	CCTGGGAGACCTGTGTGACCGACTACCTGAGAGCGCACTTGATCCCGAGTGGAGTTCCG	1909
Db	4815	CTTGGGAGGCGCTGTGTGACCGAGTACTTGGCAGGCGCACTTGATCCCGAGTGGAGTTCCG	4874
QY	1910	TGAACACCCCCCTTGGTGAAGCTTGTGTACCAAGCTGAGAGAGAGCCCATCATTCGGCG	1969
Db	4875	TGAACACCCCCCTTGGTGAAGCTTGTGTACCAAGCTGAGAGAGAGCCCATCATTCGGCG	4934
QY	1970	CCGAGACCTTCTACGTGAGACGGCGCCGCAACCGCGAGACCAAGATCTGGCAGAGCCGGCT	2029
Db	4935	CCGAGACCTTCTACGTGAGACGGCGCCGCAACCGCGAGACCAAGTCTGGCAGAGCCGGCT	4994
QY	2030	ACGTGACCCGACCGGGGCGCGAGAAAGTCTGAGCCTTGAACCGAGACCAACCAAGAGA	2089
Db	4995	ACGTGACCCGACCGGGGCGCGAGAAAGTGTGGCCCTTGAACCGACCAACCAAGAGA	5054
QY	2090	CCGAGCTGCAGGCCATCAAGTGGCCCTTGCAGGACACCGCGACGAGAGTGAACATTCCTGA	2149
Db	5055	CCGAGCTGCAGGCCATCAAGTGGCCCTTGCAGGACACCGCGCTGAGAGTGAACATTCCTGA	5114
QY	2150	CCGACAGCAGTACGCCCTTGGGCAATCAATCAGGCGCCAGCCCGACAAAGCAGAGCGAGC	2209
Db	5115	CCGACAGCAGTACGCCCTTGGGCAATCAATCAGGCGCCAGCCCGACAAAGCAGAGCGAGC	5174
QY	2210	TGTGTGAACCAAGATCATGAGACGCTGTATCAAGAAAGAGAGGTACTCTGACCTGGGTGC	2269
Db	5175	TGTGTGTGAACCAAGATCATGAGACGCTGTATCAAGAAAGAGAGGTACTCTGACCTGGGTGC	5234

DB 4583 TTTTCTTACGATCCCGACGAGGACCTGATGCGGAGATCCAGAGCAAGGCGCAGGCGCAGT 4642
QY 1670 GGAACCTACAGATCTTACAGAGAGCCCTTCAAGACCTGAGACCGGCAAGTACGCGCAAGA 1729
DB 4643 GGAACCTACAGATCTTACAGAGAGCCCTTCAAGACCTGAGACCGGCAAGTACGCGCGCA 4702
QY 1730 TGGCGACCGGCCCAACCAACGAGCTGAGAGAGCTGACCGAGGCGGTGCAAGAGATCGCCA 1789
DB 4703 TGAAGGGCGGCCCAACCAACGAGCTGAGAGAGCTGACCGAGGCGGTGCAAGAGATCGCCA 4762
QY 1790 TGGAGAGCATGGTATCTGGGGCAAGACCCCAAGTTCCGCTGCCATCCAGAGAGAGA 1849
DB 4763 CCGAGAGCATCGTATCTGGGGCAAGACCCCAAGTTCAAGCTGCCATCCAGAGAGAGA 4822
QY 1850 CCTGGAGAGACTGTGAGACCGACCTACTGACAGGCGCACTTGATCCCGAGTGGAGATTG 1909
DB 4823 CCTGGAGAGGCTGTGTGAGACCGAGTACTGACAGGCGCACTTGATCCCGAGTGGAGATTG 4882
QY 1910 TGAACACCCCGCCCTGTGTGAAGCTGTGTATCCAGCTGAGAGAGAGCCCATCATCGCG 1969
DB 4883 TGAACACCCCGCCCTGTGTGAAGCTGTGTATCCAGCTGAGAGAGAGCCCATCATCGCG 4942
QY 1970 CCGAGACCTTTACGCTGAGACCGCGCGCGCAACCGGAGACCAAGATCGGCAAGCGCGT 2029
DB 4943 CCGAGACCTTTACGCTGAGACCGCGCGCGCAACCGGAGACCAAGCTGGGCAAGCGCGT 5002
QY 2030 ACGTGAACGACCGGGGCGCGAGAGAGATGATGAGCTGACCGAGACCAACCAACAGAGA 2089
DB 5003 ACGTGAACGACCGGGGCGCGAGAGAGTGTGCGCTGACCGACCAACCAACAGAGA 5062
QY 2090 CCGAGCTGAGGCGCATCCAGCTGGCCCTGACAGGACGCGGCAAGAGGTGAACATCGTGA 2149
DB 5063 CCGAGCTGAGGCGCATCCAGCTGGCCCTGACAGGACGCGGCGCTGAGGTGAACATCGTGA 5122
QY 2150 CCGAGCGCCAGTACCGGCGCTGGGATGATCAAGGCCAGGCCCGGCAAGAGCGAGCGAGC 2209
DB 5123 CCGAGCGCCAGTACCGGCGCTGGGATGATCAAGGCCAGGCCCGGCAAGAGCGAGCGAGC 5182
QY 2210 TGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGC 2269
DB 5183 TGGTGAACAGATCATCGAGAGCTGATCAAGAGAGAGAGTGTACTGAGCTGGGTGC 5242
QY 2270 CCGGCCCAAGGGGCGATCGGCGGCAACGAGCAGATCGAACAAGCTGTGAGCAAGGCGATCC 2329
DB 5243 CCGGCCCAAGGGGCGATCGGCGGCAACGAGCAGATCGGCGGCGATCGGCGGCGATCC 5302
QY 2330 GCAAGGTGCTGTTCTGAGAGGCGATGATGGCGGCGATCGTATCTTACAGTA 2381
DB 5303 GCAAGGTGCTGTTCTGAGAGGCGATGATGGCGGCGATCGTATCTTACAGTA 5354

Search completed: December 30, 2005, 19:16:36
Job time : 11936.2 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:40 ; Search time 1293.51 Seconds
(without alignments)
12659.489 Million cell updates/sec

Title: US-09-610-313B-32

Perfect score: 2457
Sequence: 1 gtcgacgcaccacatgcccga.....gggtacgaccggtgaattc 2457

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*

- 1: geneseqn1980s:*
- 2: geneseqn1990s:*
- 3: geneseqn2000s:*
- 4: geneseqn2001as:*
- 5: geneseqn2001bs:*
- 6: geneseqn2002as:*
- 7: geneseqn2002bs:*
- 8: geneseqn2003as:*
- 9: geneseqn2003bs:*
- 10: geneseqn2003cs:*
- 11: geneseqn2003ds:*
- 12: geneseqn2004as:*
- 13: geneseqn2004bs:*
- 14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2457	100.0	2457	6	ABL39961 Synthetic
2	2457	100.0	2457	12	Adm73766 HIV-1 pol
3	2443.4	99.4	2445	8	ACA03546 Synthetic
4	2443.4	99.4	2445	10	ADCl3264 DNA of HI
5	2436.4	99.2	2463	6	ABL39960 Synthetic
6	2436.2	99.2	2463	12	Adm73765 HIV-1 pol
7	2436.2	99.1	3930	10	ADCl3231 DNA of HI
8	2434.8	99.1	3930	10	ADCl3232 DNA of HI
9	2434.8	99.1	5184	8	ACA03591 Synthetic
10	2434.8	99.1	5184	8	ACA03591 Synthetic
11	2434.8	98.8	2457	8	ACA03547 Synthetic
12	2428.6	98.8	2457	10	ADCl3265 DNA of HI
13	2428.6	98.3	2469	6	ABL39959 Synthetic
14	2415.4	98.3	2469	12	Adm73764 HIV-1 pol
15	2404.4	97.9	3531	10	ADCl3234 DNA of HI
16	2404.4	97.8	2457	8	ACA03548 Synthetic
17	2403.4	97.8	2457	10	ADCl3266 DNA of HI
18	2403.4	97.8	3537	10	ADCl3236 DNA of HI
19	2402.8	97.8	3537	10	ADCl3236 DNA of HI

20	2401.8	97.8	5145	8	ACA03521 Synthetic
21	2401.8	97.8	5145	10	ADCl3233 DNA of HI
22	2391.8	97.3	3538	10	ADCl3235 DNA of HI
23	2343.4	95.4	3607	8	ACA03551 Synthetic
24	2343.4	95.4	3607	10	ADCl3269 DNA of HI
25	2325.2	94.6	3597	8	ACA03549 Synthetic
26	2325.2	94.6	3597	10	ADCl3267 DNA of HI
27	2295.8	93.4	3624	8	ACA03550 Synthetic
28	2295.8	93.4	3624	10	ADCl3268 DNA of HI
29	2136	86.9	2460	8	ACA03541 Synthetic
30	2136	86.9	2460	8	ACC78505 HIV p2Pol
31	2135	86.9	3564	8	ACC78488 HIV GagPo
32	2135	86.9	3564	8	ACC78489 HIV GagPo
33	2134.4	86.9	4716	8	ACA03522 Synthetic
34	2134.4	86.9	4716	10	ADCl3238 DNA of HI
35	2130.6	86.7	3999	8	ACC78484 HIV GagCo
36	2129	86.7	3999	8	ACC78485 HIV GagCo
37	2129	86.7	3999	8	ACC78486 HIV GagCo
38	2129	86.7	5283	8	ACA03584 Synthetic
39	2129	86.7	5283	8	ACC78529 HIV TatRe
40	2128.8	86.6	4713	8	ACA03592 Synthetic
41	2128.8	86.6	4713	10	ADCl3280 DNA of HI
42	2128.4	86.6	3462	10	ADCl3237 DNA of HI
43	2115.2	86.1	2466	8	ACA03542 Synthetic
44	2115.2	86.1	2466	8	ACC78506 HIV p2Pol
45	2103	85.6	3735	8	ACA03545 Synthetic

ALIGNMENTS

RESULT 1	ABL39961	standard; DNA; 2457 BP.
ID	ABL39961	standard; DNA; 2457 BP.
XX	ABL39961;	
AC	ABL39961;	
XX	15-MAY-2002 (first entry)	
DT	Synthetic construct PR97SYMM SEQ ID NO:32.	
XX	Human immunodeficiency virus type C; antigenic HIV type C protein;	
KW	immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;	
KW	immunostimulant; gene therapy; gene; ds.	
XX	Human immunodeficiency virus; type C.	
OS	Synthetic.	
XX	WO200204493-A2.	
PN	17-JAN-2002.	
XX	05-JUL-2001; 2001WO-US021241.	
PF	05-JUL-2000; 2000US-00610313.	
PR	(CHIR) CHIRON CORP.	
PA	(UYST-) UNIV STELLENBOSCH.	
XX	Zur Megede J, Barnett SW, Engelbrecht S, Van Rensburg EJ;	
PI	WPI, 2002-154920/20.	
DR	New polynucleotides encoding antigenic HIV Type C polypeptides, useful in	
XX	applications including DNA immunization or generation of packaging cell	
PT	lines, particularly in gene therapy.	
PT	Claim 1; Fig 10; 233pp; English.	
XX	The present invention describes expression cassettes comprising a	
PS	polynucleotide sequence encoding a polypeptide comprising immunogenic HIV	
CC	type C polypeptides. The expression cassettes comprise any of the HIV	
CC	type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef	

CC (1). (1) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunization. Generation of packaging cell lines, and production of HIV
CC Type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABLJ9942 to ABL40054 and
CC ABL06204 to ABL06215 represent sequences used in the exemplification of
CC the present invention
CC XX

SQ Sequence 2457 BP; 566 A; 837 C; 754 G; 300 T; 0 U; 0 Other;

Query Match 100.0%; Score 2457; DB 6; Length 2457;

Best Local Similarity 100.0%; Pred. No. 1,1e-295;

Matches 2457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 GTGAGGCGCAACCATGAGGCGCATGAGCCGAGGCGCACAGGCGCAACCTCTGATGAG 60
DB 1 GTGAGGCGCAACCATGAGGCGCATGAGCCGAGGCGCACAGGCGCAACCTCTGATGAG 60
QY 61 CGAGCAACTTCAAGGGGCGCCAGCGCATCATCAAGTCTTCAACTGCGCGCAAGAGGAGC 120
DB 61 CGAGCAACTTCAAGGGGCGCCAGCGCATCATCAAGTCTTCAACTGCGCGCAAGAGGAGC 120
QY 121 CACATCGCCCGCACTGCGCGCCCGCGCGCAAGAGGCGTGTGGAAGTGCAGAGAG 180
DB 121 CACATCGCCCGCACTGCGCGCCCGCGCGCAAGAGGCGTGTGGAAGTGCAGAGAG 180
QY 181 GCGCACAAGATGAAGACTGCAACGAGGCGCAGGCGCAACTTCTTCCGAGAGGACTGAGC 240
DB 181 GCGCACAAGATGAAGACTGCAACGAGGCGCAGGCGCAACTTCTTCCGAGAGGACTGAGC 240
QY 241 TTCCCGCAAGGCGAAGGCGCGAGTTCCCGCAGCGAGAGAGACCGGCGCCAGGCCCAAC 300
DB 241 TTCCCGCAAGGCGAAGGCGCGAGTTCCCGCAGCGAGAGAGACCGGCGCCAGGCCCAAC 300
QY 301 AGCGCGAGAGCTGCAAGTGCAGGCGCAACACCCCGCGAGGCGCGGCGCGAGCGCGAG 360
DB 301 AGCGCGAGAGCTGCAAGTGCAGGCGCAACACCCCGCGAGGCGCGGCGCGAGCGCGAG 360
QY 361 GGCACCTGGAATCTTCCCGCAAGTCAACCTGTGAGAGCGCCCTGTGAGCATCAAGGAG 420
DB 361 GGCACCTGGAATCTTCCCGCAAGTCAACCTGTGAGAGCGCCCTGTGAGCATCAAGGAG 420
QY 421 GCGCGCGCAAGATCAAGAGAGGCGCTGTGAGCAACGCGCGCGAGCAACGCGTGTGAGAG 480
DB 421 GCGCGCGCAAGATCAAGAGAGGCGCTGTGAGCAACGCGCGCGAGCAACGCGTGTGAGAG 480
QY 481 ATGAGCTGCGCGCAAGTGAAGAGCCCAAGATGATCGCGGCGCATCGCGGCTTCACTAAG 540
DB 481 ATGAGCTGCGCGCAAGTGAAGAGCCCAAGATGATCGCGGCGCATCGCGGCTTCACTAAG 540
QY 541 GTGCGCGAGTACAGACAGATCTGTATCGAGATCTGCGGCGAAGAGGCGCATCGGCGAG 600
DB 541 GTGCGCGAGTACAGACAGATCTGTATCGAGATCTGCGGCGAAGAGGCGCATCGGCGAG 600
QY 601 CTGATCGGCGCGCAACCCCGTGAACATCATCGGCGCGCAACATGCTGAGCTGAGCTGC 660
DB 601 CTGATCGGCGCGCAACCCCGTGAACATCATCGGCGCGCAACATGCTGAGCTGAGCTGC 660
QY 661 ACCCTGAACCTTCCCATGAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATG 720
DB 661 ACCCTGAACCTTCCCATGAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATG 720
QY 721 GACGCGCCCAAGAGTGAAGAGTGCCTGTGAACGAGAGAGAGATGAAGGCGCTTGACCGGC 780
DB 721 GACGCGCCCAAGAGTGAAGAGTGCCTGTGAACGAGAGAGAGATGAAGGCGCTTGACCGGC 780
QY 781 ATCTCGAGAGAGATGAAGAGAGGCGCAAGATCAACCAAGATCGGCGCGCGAGACCTTAC 840
DB 781 ATCTCGAGAGAGATGAAGAGAGGCGCAAGATCAACCAAGATCGGCGCGCGAGACCTTAC 840
QY 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGCAGAGCTGTGTGAGC 900
DB 841 AACACCCCGGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGCAGAGCTGTGTGAGC 900
```

```
QY 901 TTCCGCGAGCTGAACAAGCGACCCAGGACTTCTGAGAGGTGAGCTGAGGATCCCCCAGC 960
DB 901 TTCCGCGAGCTGAACAAGCGACCCAGGACTTCTGAGAGGTGAGCTGAGGATCCCCCAGC 960
QY 961 CCCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGACGTGAGCGAGCGCTTACTTC 1020
DB 961 CCCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGACGTGAGCGAGCGCTTACTTC 1020
QY 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACATCCAGCATCAAC 1080
DB 1021 AGCGTGCCTTGAAGAGAGACTTCCGCAAGTACACCGGCTTCAACATCCAGCATCAAC 1080
QY 1081 AACGAGACCCCGGCGCATCGCTACCAAGTCAACGTCGCGCCAGGCGTGAAGAGGAGC 1140
DB 1081 AACGAGACCCCGGCGCATCGCTACCAAGTCAACGTCGCGCCAGGCGTGAAGAGGAGC 1140
QY 1141 CCGAGCATCTTCAAGAGAGATGAACCAAGTCTGTAGAGCCCTTCCGCGCCGCAACCCC 1200
DB 1141 CCGAGCATCTTCAAGAGAGATGAACCAAGTCTGTAGAGCCCTTCCGCGCCGCAACCCC 1200
QY 1201 GAGATCGTGAATCAAGAGCCCGCTGTACGTGAGCAAGCGACCTGAGAGATCGAGCAGC 1260
DB 1201 GAGATCGTGAATCAAGAGCCCGCTGTACGTGAGCAAGCGACCTGAGAGATCGAGCAGC 1260
QY 1261 GCGCGCAAGATCGAGAGCTGCGCAAGACCTGTGCGCTGAGGCGTTCAACACCCCGAGC 1320
DB 1261 GCGCGCAAGATCGAGAGCTGCGCAAGACCTGTGCGCTGAGGCGTTCAACACCCCGAGC 1320
QY 1321 AAGAGCAACGAAGAGAGAGCCCTTCTGCGCATGAGTGCACCCCGCAAGTGAAC 1380
DB 1321 AAGAGCAACGAAGAGAGAGCCCTTCTGCGCATGAGTGCACCCCGCAAGTGAAC 1380
QY 1381 GTGAGAGCCCATGAGCTGCGCGAGAGAGAGCTGGAACCGTGAACGATCAAGAGAGCTG 1440
DB 1381 GTGAGAGCCCATGAGCTGCGCGAGAGAGAGCTGGAACCGTGAACGATCAAGAGAGCTG 1440
QY 1441 GTGAGAGAGCTGAATCTGAGGCGAGGAGATCAACCCCGCATCAAGTGCAGGCTGAGC 1500
DB 1441 GTGAGAGAGCTGAATCTGAGGCGAGGAGATCAACCCCGCATCAAGTGCAGGCTGAGC 1500
QY 1501 AAGCTGCTGCGCGCGCAAGGCGCTTGAACGAGATGTGCTTGAACCGAGAGGCGAG 1560
DB 1501 AAGCTGCTGCGCGCGCGCAAGGCGCTTGAACGAGATGTGCTTGAACCGAGAGGCGAG 1560
QY 1561 CTGAGAGCTGCGCGAGAGACCGGAGAGTCTGTGCGCGAGCCCGTGCACGCGGTGAACAG 1620
DB 1561 CTGAGAGCTGCGCGAGAGACCGGAGAGTCTGTGCGCGAGCCCGTGCACGCGGTGAACAG 1620
QY 1621 CCGAGCAAGAGACTGTGAGCGAGATCCAGAGACGAGGCGACGACCAAGTGAACCTAACG 1680
DB 1621 CCGAGCAAGAGACTGTGAGCGAGATCCAGAGAGCGAGGCGACGACCAAGTGAACCTAACG 1680
QY 1681 ATTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGAGCACGCGC 1740
DB 1681 ATTACCAAGAGAGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCAAGATGAGCACGCGC 1740
QY 1741 CACACCAACAGAGTGAAGACGTGACCGAGGCGGTGCGAGAGATGCGCATGAGAGAGCTC 1800
DB 1741 CACACCAACAGAGTGAAGACGTGACCGAGGCGGTGCGAGAGATGCGCATGAGAGAGCTC 1800
QY 1801 GTGATCTGAGGAGAGAGAGAGAGAGTTCGCGCTGAGGCGCATCAAGAGAGAGAGCTGAGAG 1860
DB 1801 GTGATCTGAGGAGAGAGAGAGAGAGTTCGCGCTGAGGCGCATCAAGAGAGAGAGCTGAGAG 1860
QY 1861 TGTGAGACCGACTATGAGAGAGCGCACTGAGATCCCGAGTGGAGATTGTGTAACACCCC 1920
DB 1861 TGTGAGACCGACTATGAGAGAGCGCACTGAGATCCCGAGTGGAGATTGTGTAACACCCC 1920
QY 1921 CCCCTGTGAGAGCTGTGTACAGCTGAGAGAGAGAGAGGCGCATCATGAGGCGCGAGACTTC 1980
DB 1921 CCCCTGTGAGAGCTGTGTGTACAGCTGAGAGAGAGAGAGGCGCATCATGAGGCGCGAGACTTC 1980
```


Db 781 ATTCGAGAGAGATGAGAGAGGCAAGATCAACAAAGATCGAGCCCGAGAACCCCTTAC 840
Qy AACACCCCGTGTGGCCATCAAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGGAC 900
Db AACACCCCGTGTGGCCATCAAGAGAGAGACAGACCAAGTGGCGCAAGCTGTGGAC 900
Qy TTCCGAGAGCTGAACAGACGCAACCGAGACTTCTGGAGGTGCACTGGGCAATCCCGAC 960
Db TTCCGAGAGCTGAACAGACGCAACCGAGACTTCTGGAGGTGCACTGGGCAATCCCGAC 960
Qy CCCGCGGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGAGCGCTTACTTC 1020
Db CCCGCGGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGGGCGAGCGCTTACTTC 1020
Qy AGCGTGGCCCTGAGAGAGAGACTTTCGCAAGTACACCGGCTTCAACCTCCAGCACTAAC 1080
Db AGCGTGGCCCTGAGAGAGAGACTTTCGCAAGTACACCGGCTTCAACCTCCAGCACTAAC 1080
Qy AACGAGACCCCGGCGATCCGCTACAGATCAACGTGTGCGCCAGAGGCTGAGAGGCGACG 1140
Db AACGAGACCCCGGCGATCCGCTACAGATCAACGTGTGCGCCAGAGGCTGAGAGGCGACG 1140
Qy CCCAGCATCTTTCAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCCGCAACCC 1200
Db CCCAGCATCTTTCAGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCCGCAACCC 1200
Qy GAGATCGATGATCAACAGAGCCCGCTGTACCGTGGGAGAGAGCTTGGAGAGTGGGCGAC 1260
Db GAGATCGATGATCAACAGAGCCCGCTGTACCGTGGGAGAGAGCTTGGAGAGTGGGCGAC 1260
Qy CGCGCAAGATGAGAGAGCTGGCAAGACCTGTGGCTGGGCGCTTCAACACCCCGAC 1320
Db CGCGCAAGATGAGAGAGCTGGCAAGACCTGTGGCTGGGCGCTTCAACACCCCGAC 1320
Qy AAGAGAGCAACAG 1380
Db AAGAGAGCAACAG 1380
Qy AAGAGAGCAACAG 1380
Db AAGAGAGCAACAG 1380
Qy GTGCAAGCCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db GTGCAAGCCCATCGAGCTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy GTGGCAAGCTGAACCTGGGCGAGAGATCTACCCCGGATCAAGTGGGCGAGCTGTGC 1500
Db GTGGCAAGCTGAACCTGGGCGAGAGATCTACCCCGGATCAAGTGGGCGAGCTGTGC 1500
Qy AAGCTGTGCGCGGCGCAAGGCGCTGACCGACATCGTGCCTTGAACCGAGAGAGCGGAG 1560
Db AAGCTGTGCGCGGCGCAAGGCGCTGACCGACATCGTGCCTTGAACCGAGAGAGCGGAG 1560
Qy CTGAGAGCTGGCGAGAGACCGAGAGATCTGTGCGGAGAGCGCGGTGTACTAGAC 1620
Db CTGAGAGCTGGCGAGAGACCGAGAGATCTGTGCGGAGAGCGCGGTGTACTAGAC 1620
Qy CCCAGCAAGAGACTGTGTGCGCGAGATCTCAAGAGAGGCGACAGCAAGTGAACCTTACG 1680
Db CCCAGCAAGAGACTGTGTGCGCGAGATCTCAAGAGAGGCGACAGCAAGTGAACCTTACG 1680
Qy ATCTACCAAGAGAGCTTTCAGAGAGCTGAAGAGCGGCAAGTACGCAAGTGGCGACCGCC 1740
Db ATCTACCAAGAGAGCTTTCAGAGAGCTGAAGAGCGGCAAGTACGCAAGTGGCGACCGCC 1740
Qy CACACCAAGAGAGCTGAAGAGAGCTGACCGAGCGGTGAGAGAGATGGCAATGGAGAGAC 1800
Db CACACCAAGAGAGCTGAAGAGAGCTGACCGAGCGGTGAGAGAGATGGCAATGGAGAGAC 1800
Qy GTGATCTGGGCGAGAGCCCGCAAGTTCGCGCTGCCATTCAGAGAGAGAGCTGGAGAGCC 1860
Db GTGATCTGGGCGAGAGCCCGCAAGTTCGCGCTGCCATTCAGAGAGAGAGCTGGAGAGCC 1860
Qy TGGTGAACCGAGATCTGGGAGAGCGACCTGATCCCGAGTGGAGATTGATGAACACCCCG 1920
Db TGGTGAACCGAGATCTGGGAGAGCGACCTGATCCCGAGTGGAGATTGATGAACACCCCG 1920

Qy 1921 CCCCTGGTGAAGCTGTGTGACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 1921 CCCCTGGTGAAGCTGTGTGACCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy 1981 TACGTGAACGAGCGCGCCCAACCGCGAGAGCAAGATTCGCAAGAGCGGCTAGGTGACGAC 2040
Db 1981 TACGTGAACGAGCGCGCCCAACCGCGAGAGCAAGATTCGCAAGAGCGGCTAGGTGACGAC 2040
Qy 2041 CGGAGCGGCGAGAGAGTGTGAGCTTGAACCGAGAGCAACCGAGAGAGAGAGAGAGAGAG 2100
Db 2041 CGGAGCGGCGAGAGAGTGTGAGCTTGAACCGAGAGCAACCGAGAGAGAGAGAGAGAGAG 2100
Qy 2101 GGCATTCAGCTGGCGCTTGAAG 2160
Db 2101 GGCATTCAGCTGGCGCTTGAAG 2160
Qy 2161 TACGCGCTGGGATCATTCAGAGCGCCAGCCGAGCAAGAGAGAGAGAGAGAGAGAGAGAG 2220
Db 2161 TACGCGCTGGGATCATTCAGAGCGCCAGCCGAGCAAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2221 ATCATTCAGAGAGCTGATTCAG 2280
Db 2221 ATCATTCAGAGAGCTGATTCAG 2280
Qy 2281 GGCATTCGAGCGGCAACAGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Db 2281 GGCATTCGAGCGGCAACAGAGAGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy 2341 TTCTTGAACGAGAGATGATGAGCGGAGATCGATGATCTTACAGATGATGATGATGATGAT 2400
Db 2341 TTCTTGAACGAGAGATGATGAGCGGAGATCGATGATCTTACAGATGATGATGATGATGAT 2400
Qy 2401 GGCAGCGGCGGCGCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2457
Db 2401 GGCAGCGGCGGCGCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2457

RESULT 3
ACA03546
ID ACA03546 standard, DNA, 2445 BP.
XX
AC ACA03546;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #29.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; de.
OS Synthetic.
XX
PN WO2003004657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002WO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX
PR 31-AUG-2001; 2001US-0316860P.
XX
PR 16-JAN-2002; 2002US-0349728P.
XX
PR 16-JAN-2002; 2002US-0349793P.
XX
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PI Zur Megede J, Barnett SM, Llan Y;
XX
DR WPI; 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C

polypeptides, useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, especially humans.

Example 1; Fig 34; 262bp; English.

The invention describes a synthetic polynucleotide encoding 2 or more immunogenic HIV polypeptides, where at least 2 of the polypeptides are derived from different HIV subtypes. The polynucleotide is useful for immunization, generation of packaging cell lines, or production of HIV polypeptides. The polynucleotide and its encoded proteins are useful as immunogenic compositions or vaccines for generating humoral or cellular immune responses against HIV in a subject, or for inducing neutralising antibodies against HIV. The gene delivery vector comprising the polynucleotide is also useful for DNA immunisation of, or for generating an immune response (e.g. a humoral or cellular immune response) in, a subject such as a mammal, particularly a human. This sequence encodes a human immunodeficiency virus immunogenic peptide

Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;

Query Match 99.4%; Score 2443.4; DB 8; Length 2445;

Best Local Similarity 100.0%; Pred. No. 5.3e-294; Mismatches 1; Indels 0; Gaps 0;

Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 7 GCCACCATGCGCGAGCCATGAGCCAGCGCCACATCTTGATGACGCGACG 66
DB 1 GCCACCATGCGCGAGCCATGAGCCAGCGCCACATCTTGATGACGCGACG 60
QY 67 AACTTCAGAGGCGCCCAAGGCGATCATCAAGTCTTCAATGCGCGCAAGAGGCGCCATC 126
DB 61 AACTTCAGAGGCGCCCAAGGCGATCATCAAGTCTTCAATGCGCGCAAGAGGCGCCATC 120
QY 127 GCCCGCAACTGCGCGCGCGCGCGCGCGCAAGAGGCGTGTGAGAGTGCGCGCAAGAGGCGCGAC 186
DB 121 GCCCGCAACTGCGCGCGCGCGCGCGCGCGCAAGAGGCGTGTGAGAGTGCGCGCAAGAGGCGCGAC 180
QY 187 CAGATGAAGAGATGCAACCGAGCGCCAGCCACTTCTTCGCGAGAGACTGTGCGCTTCC 246
DB 181 CAGATGAAGAGATGCAACCGAGCGCCAGCCACTTCTTCGCGAGAGACTGTGCGCTTCC 240
QY 247 CAGGCGAGAGGCGCGCGAGTTCCTCCAGCGAGCAAGACCGCGCCACAGCGCGCGCGCGCG 306
DB 241 CAGGCGAGAGGCGCGCGAGTTCCTCCAGCGAGCAAGACCGCGCCACAGCGCGCGCGCGCG 300
QY 307 GAGCTGAGAGTGCAGCGCGCGCAACCCCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 366
DB 301 GAGCTGAGAGTGCAGCGCGCGCAACCCCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 360
QY 367 CTGAACCTTCCCGCAGATCAACCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGCGCGCG 426
DB 361 CTGAACCTTCCCGCAGATCAACCTGTGCGAGCGCGCGCTGTGAGCATCAAGGTGCGCGCG 420
QY 427 CAGATCAAGAGAGGCGCTGTGAGCACCGCGCGCGCGAGCAACCGTGTGAGAGAGATGAC 486
DB 421 CAGATCAAGAGAGGCGCTGTGAGCACCGCGCGCGCGAGCAACCGTGTGAGAGAGATGAC 480
QY 487 CTGCGCGCGAGAGTGAAGCGCGCAAGATGATCGCGCGAGATCGCGCGCTTCAAGAGTGC 546
DB 481 CTGCGCGCGAGAGTGAAGCGCGCAAGATGATCGCGCGAGATCGCGCGCTTCAAGAGTGC 540
QY 547 CAGTACGACGAGATCTGTATCGAGATCTGCGCGCAAGAGCGCATCGGCAACCGTGTGATC 606
DB 541 CAGTACGACGAGATCTGTATCGAGATCTGCGCGCAAGAGCGCATCGGCAACCGTGTGATC 600
QY 607 GCGCGCGCGCGCGTGAACATCATCGCGCGCGCAAGTCTGACCGCTGCGCGCGCGCGCG 666
DB 601 GCGCGCGCGCGCGTGAACATCATCGCGCGCGCAAGTCTGACCGCTGCGCGCGCGCGCG 660
QY 667 AACTTCCCGCATGAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCGCATGAGCG 726
DB 661 AACTTCCCGCATGAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCGCATGAGCG 720
```

```

QY 727 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGGCATCTGC 786
DB 721 CCAAGGTGAAGCAGTGGCGCCCTGACCGAGGAGAGATCAAGGCGCTGACCGGCATCTGC 780
QY 787 GAGGAGATGAGAGAGAGGCGCAAGATCAACAGATCGCGCGCGAGAACCCCTTACAACAC 846
DB 781 GAGGAGATGAGAGAGAGGCGCAAGATCAACAGATCGCGCGCGAGAACCCCTTACAACAC 840
QY 847 CCGGTGTTCCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
DB 841 CCGGTGTTCCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
QY 907 GAGCTGAACAGCGCACCGAGACTTCTGAGAGTGAAGTGTGAGAGTGTGAGAGTGTGAGAG 966
DB 901 GAGCTGAACAGCGCACCGAGACTTCTGAGAGTGAAGTGTGAGAGTGTGAGAGTGTGAGAG 960
QY 967 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1026
DB 961 GCGCTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
QY 1027 CCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
DB 1021 CCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1087 ACCCGCGCGATCCGCTTACAGTGAACAGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1146
DB 1081 ACCCGCGCGATCCGCTTACAGTGAACAGTGTGCGCGCGCGCGCGCGCGCGCGCGCG 1140
QY 1147 ATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
DB 1141 ATCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
QY 1207 GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
DB 1201 GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
QY 1267 AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
DB 1261 AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
QY 1327 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
DB 1321 CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1387 CCGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
DB 1381 CCGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1447 AAGCTGAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
DB 1441 AAGCTGAACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1507 CTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1566
DB 1501 CTGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1567 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1626
DB 1561 CTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1627 AAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
DB 1621 AAGAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1687 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
DB 1681 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
QY 1747 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806
DB 1741 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1807 TGGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
```

Db 1801 TGGGGCAGAGCCCGCAAGTTCGCTGCCATCCAGAGAGAGACTGGGAGACTGGTGG 1860
QY 1867 ACCGACTCTGTCAGAGCCGACCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCTGG 1926
Db 1861 ACCGACTCTGTCAGAGCCGACCTGGATCCCGAGTGGGAGTTCTGTAAACACCCCTGG 1920
QY 1927 GTGAAGCTGTGTGTCACAGCTGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTC 1986
Db 1921 GTGAAGCTGTGTGTCACAGCTGAGAGAGAGCCCATCATCGGCGCGAGACCTTCTACGTC 1980
QY 1987 GACGCGCCCGCCCAACCGGAGAGCCAGATCGGCAAGCCCGCTACGTGACCGAGCGGC 2046
Db 1981 GACGCGCCCGCCCAACCGGAGAGCCAGATCGGCAAGCCCGCTACGTGACCGAGCGGC 2040
QY 2047 CGGCAAGATCTGTGTCAGCTGAGAGAGAGCCAGCCAGCCAGAGAGACCGAGCTGAGCGCATC 2106
Db 2041 CGGCAAGATCTGTGTCAGCTGAGAGAGAGCCAGCCAGCCAGAGAGACCGAGCTGAGCGCATC 2100
QY 2107 CAGCTGGCCCTGTCAGAGAGCGGAGAGGTGAACATCTGTGACCGAGCGAGCTGACG 2166
Db 2101 CAGCTGGCCCTGTCAGAGAGCGGAGAGGTGAACATCTGTGACCGAGCGAGCTGACG 2160
QY 2167 CTGGGCACTATCCAGAGCCCGACCGGCAAGAGAGAGCGAGCTGTGAAACCAATCATC 2226
Db 2161 CTGGGCACTATCCAGAGCCCGACCGGCAAGAGAGAGCGAGCTGTGAAACCAATCATC 2220
QY 2227 GAGCGCTGATCAAAAGAGAGAGAGTGTACCTGAGACTGGTCCCGCCCAAGGGCATC 2286
Db 2221 GAGCGCTGATCAAAAGAGAGAGAGTGTACCTGAGACTGGTCCCGCCCAAGGGCATC 2280
QY 2287 GCGCGCAAGAGAGATCGACAGAGCTGTGTAGAGAGGAGCATCCGAAAGTGTCTTCTG 2346
Db 2281 GCGCGCAAGAGAGATCGACAGAGCTGTGTAGAGAGGAGCATCCGAAAGTGTCTTCTG 2340
QY 2347 GACGCGCATGATGCGGAGCTGTGTATCTACAGTACATGAGAGACCTGTGACGTGGGAGC 2406
Db 2341 GACGCGCATGATGCGGAGCTGTGTATCTACAGTACATGAGAGACCTGTGACGTGGGAGC 2400
QY 2407 GCGCGCCCTTGAAGATGATTTAAAGCTTCCCGGGGCTTACGACCGGT 2451
Db 2401 GCGCGCCCTTGAAGATGATTTAAAGCTTCCCGGGGCTTACGACCGGT 2445

RESULT 4
ADCl3264
ID ADCl3264 standard; DNA, 2445 BP.
XX
AC ADCl3264;
XX
DT 18-DEC-2003 (first entry)
XX
DB DNA of HIV construct p2Pol-opt-YMMW_C SEQ ID NO 43.
XX
KW expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prox;
XX
OS Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; de.
XX
OS Human immunodeficiency virus.
XX
PN MO2003004620-A2.
XX
PD 16-JAN-2003.
XX
PE 05-JUL-2002; 2002MO-US021420.
XX
PR 05-JUL-2001; 2001US-0303192P.
XX
PR 31-AUG-2001; 2001US-0316860P.
XX
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR) CHIRON CORP.
XX
PA (UYST-) UNIV STELLENBOSCH.
PI Zur Megele J, Barnett SM, Llan Y, Engelbrecht S, Van Renburg Et;

XX
DR WPI, 2003-221593/21.
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV gag, Env, Int, Nef, p15RNaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
PS Disclosure: Fig 40; 301dp; English.
XX
CC The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV gag, Env,
CC Int, Nef, p15RNaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
SQ Sequence 2445 BP; 562 A; 835 C; 751 G; 297 T; 0 U; 0 Other;
Query Match 99.4%; Score 2443.4; DB 10; Length 2445;
Best Local Similarity 100.0%; Pred. No. 5.3e-294;
Matches 2444; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 7 GCCACATGCGCCGAGGCGCATGAGCCAGCCAGCCGCAACATCTGATGACGCGCAGC 66
Db 1 GCCACATGCGCCGAGGCGCATGAGCCAGCCAGCCGCAACATCTGATGACGCGCAGC 60
QY 67 AACTTCAAGGGGCCCCAAGCGATCATCAAGTCTTCACTGCGGCAAGAGGGCCACATC 126
Db 61 AACTTCAAGGGGCCCCAAGCGATCATCAAGTCTTCACTGCGGCAAGAGGGCCACATC 120
QY 127 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGAGGCTGCTGAAAGTGGCGCAAGAGGGCCAC 186
Db 121 GCCCGCAACTGCGCGCGCCCGCCGCAAGAGAGGCTGCTGAAAGTGGCGCAAGAGGGCCAC 180
QY 187 CAGATGAAGAGCTGACCGAGGCGCCAGGCAACTTCTTCCGAGAGACTTGGCTTCC 246
Db 181 CAGATGAAGAGCTGACCGAGGCGCCAGGCAACTTCTTCCGAGAGACTTGGCTTCC 240
QY 247 CAGGCGCAAGGCGCGGAGATTCGCCAGGAGAGAGAGCGGCAAGGCCAGCGCGC 306
Db 241 CAGGCGCAAGGCGCGGAGATTCGCCAGGAGAGAGAGCGGCAAGGCCAGCGCGC 300
QY 307 GAGCTGAGAGTGCAGGCGGAGCAACCCCGCAGAGAGCGCGCGCGCGCGCGCGCGCGC 366
Db 301 GAGCTGAGAGTGCAGGCGGAGCAACCCCGCAGAGAGCGCGCGCGCGCGCGCGCGCGC 360
QY 367 CTGAACCTTCCCGCAGATCACCTGTGTGAGAGGCGCCCTGTGTGAGATCAAGTGGCGGC 426
Db 361 CTGAACCTTCCCGCAGATCACCTGTGTGAGAGGCGCCCTGTGTGAGATCAAGTGGCGGC 420
QY 427 CAGATCAAGAGAGGCTGTGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGC 486
Db 421 CAGATCAAGAGAGGCTGTGTGTGAGCAACCGGCGCGCGCGCGCGCGCGCGCGCGCGC 480
QY 487 CTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGAGCATCGGCGGCTTATCAACAGTGGCGC 546
Db 481 CTGCGCGGCAAGTGAAGCCCAAGATGATCGGCGGAGCATCGGCGGCTTATCAACAGTGGCGC 540
QY 547 CAGTACGACAGATCTGATGAGATCTGCGGCAAGAGGCGCATCGGCGCGTGTGATC 606
Db 541 CAGTACGACAGATCTGATGAGATCTGCGGCAAGAGGCGCATCGGCGCGTGTGATC 600
QY 607 GCGCCCAACCCCGTGAATCATGAGCGGCGCAACATGCTGAGCCAGCTGGGCTGACCCCTG 666
Db 601 GCGCCCAACCCCGTGAATCATGAGCGGCGCAACATGCTGAGCCAGCTGGGCTGACCCCTG 660
QY 667 AACTTCCCATGAGCCCATGAGACGCTGCGCGTGAAGTGAAGCCCGGCGCATGAGCGGC 726

```

Db      661  |||||
      661  AACTTCCCATCAGCCCATCGAGACCTGTCCTGAAAGCTGAAGCCCGGCAATGACCGC 720
Qy      727  |||||
      727  CCGAAGGTGAAGCACTGAGCCCTGACCGAGAAAGATCAAGGCCCTGACCGCATCTGC 786
Db      721  |||||
      721  CCGAAGGTGAAGCACTGAGCCCTGACCGAGAAAGATCAAGGCCCTGACCGCATCTGC 780
Qy      787  |||||
      787  GAGGAGATGAGAAAGAGAGGCAAGATCACCAAGATCGGCCCGGAAACCCTTCAACAACC 846
Db      781  |||||
      781  GAGGAGATGAGAAAGAGAGGCAAGATCACCAAGATCGGCCCGGAAACCCTTCAACAACC 840
Qy      847  |||||
      847  CCGGTTTCGACCATCAAGAAAGAGACAGCAACAGTGGCGGAGCTGTGTGACTTCCGC 906
Db      841  |||||
      841  CCGGTTTCGACCATCAAGAAAGAGACAGCAACAGTGGCGGAGCTGTGTGACTTCCGC 900
Qy      907  |||||
      907  GAGCTGAACAAGCGCAACCGAGACTTCTGGAGGTGACAGCTGGGATCCCGCACCGCGC 966
Db      901  |||||
      901  GAGCTGAACAAGCGCAACCGAGACTTCTGGAGGTGACAGCTGGGATCCCGCACCGCGC 960
Qy      967  |||||
      967  GAGCTGAAGAAAGAAAGAGAGCTGACCGTGTGACGTGGCGGAGCGCTTACGCTG 1026
Db      961  |||||
      961  GAGCTGAAGAAAGAAAGAGAGCTGACCGTGTGACGTGGCGGAGCGCTTACGCTG 1020
Qy      1027  |||||
      1027  CCGCTGAGAGAGAGACTTCCGCAAGTACACCGCTTCAACCATCCGAGATCAACAAGAG 1086
Db      1021  |||||
      1021  CCGCTGAGAGAGAGACTTCCGCAAGTACACCGCTTCAACCATCCGAGATCAACAAGAG 1080
Qy      1087  |||||
      1087  ACCCGCGGATCCGCTACCAAGTACAAAGTCTGACCCGAGGGCTGAAAGGGACCGCAC 1146
Db      1081  |||||
      1081  ACCCGCGGATCCGCTACCAAGTACAAAGTCTGACCCGAGGGCTGAAAGGGACCGCAC 1140
Qy      1147  |||||
      1147  ATCTTTCAGAGAGAGATGACCAAGATCTGAGAGCTTTCGCGCGCCGCAACCCAGATC 1206
Db      1141  |||||
      1141  ATCTTTCAGAGAGAGATGACCAAGATCTGAGAGCTTTCGCGCGCCGCAACCCAGATC 1200
Qy      1207  |||||
      1207  GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1266
Db      1201  |||||
      1201  GTGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
Qy      1267  |||||
      1267  AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1326
Db      1261  |||||
      1261  AAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1320
Qy      1337  |||||
      1337  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1386
Db      1321  |||||
      1321  CACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
Qy      1387  |||||
      1387  CCGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1446
Db      1381  |||||
      1381  CCGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Qy      1447  |||||
      1447  AAGCTGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1506
Db      1441  |||||
      1441  AAGCTGAATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Qy      1507  |||||
      1507  CTGCGGCGGCGCAAGGCGCTGACCGACATCTGCGCCCTGACCGAGAGAGCGAGCTGAG 1566
Db      1501  |||||
      1501  CTGCGGCGGCGCAAGGCGCTGACCGACATCTGCGCCCTGACCGAGAGAGCGAGCTGAG 1560
Qy      1567  |||||
      1567  CTGCGCGAGAGACCGCGAGATCTCTGCGCGAGAGCGCGGTGTATCTACGAGCCGAC 1626
Db      1561  |||||
      1561  CTGCGCGAGAGACCGCGAGATCTCTGCGCGAGAGCGCGGTGTATCTACGAGCCGAC 1620
Qy      1627  |||||
      1627  AAGGACCTGTGGCGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1686
Db      1621  |||||
      1621  AAGGACCTGTGGCGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Qy      1687  |||||
      1687  CAGGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1746
Db      1681  |||||
      1681  CAGGAGCCCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
Qy      1747  |||||
      1747  AACGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1806

```

```

Db      1741  |||||
      1741  AACGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Qy      1807  |||||
      1807  TGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1866
Db      1801  |||||
      1801  TGGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Qy      1867  |||||
      1867  ACCGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1926
Db      1861  |||||
      1861  ACCGACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy      1927  |||||
      1927  GTGAAGCTGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1986
Db      1921  |||||
      1921  GTGAAGCTGTGTGATCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Qy      1987  |||||
      1987  GACGCGCGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2046
Db      1981  |||||
      1981  GACGCGCGCGCAACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2040
Qy      2047  |||||
      2047  CCGGCAAGAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2106
Db      2041  |||||
      2041  CCGGCAAGAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2100
Qy      2107  |||||
      2107  CAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2166
Db      2101  |||||
      2101  CAGCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2160
Qy      2167  |||||
      2167  CTGGGCAATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2226
Db      2161  |||||
      2161  CTGGGCAATCATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy      2227  |||||
      2227  GAGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2286
Db      2221  |||||
      2221  GAGCAGCTGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2280
Qy      2287  |||||
      2287  GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2346
Db      2281  |||||
      2281  GCGGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2340
Qy      2347  |||||
      2347  GACGCAATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2406
Db      2341  |||||
      2341  GACGCAATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2400
Qy      2407  |||||
      2407  GCGGCGCTTGAAGATGATTTAAAGCTTCCCGGGGCTTGAAGAGAGAGAGAGAGAG 2451
Db      2401  |||||
      2401  GCGGCGCTTGAAGATGATTTAAAGCTTCCCGGGGCTTGAAGAGAGAGAGAGAGAG 2445

RESULT 5
ADCI3230
ID      ADCI3230 standard; DNA; 3930 BP.
XX
AC      ADCI3230;
XX
XX
DT      18-DEC-2003 (first entry)
XX
DE      DNA of HIV construct GagComp]Polmut_C SEQ ID NO 9.
XX
KW      expression cassette; HIV Gag; Env; Int; Nef; p15RNaseH; Pol; Tat; Prot;
XX
OS      Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX
XX      Human immunodeficiency virus.
XX
XX      WO2003004620-A2.
XX
XX      16-JAN-2003.
XX
XX      05-JUL-2002; 2002WO-US021420.
XX
XX      05-JUL-2001; 2001US-0303192P.
XX
XX      31-AUG-2001; 2001US-031860P.
XX
XX      16-JAN-2002; 2002US-0349871P.
XX
XX      (CHIR ) CHIRON CORP.

```

PA (U957-1) UNIV STELLENBOSCH.
XX Zur Megede J, Barnett SM, Lian Y, Engelbrecht S, Van Rensburg BJ;
XX WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tct,
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 6; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intradermally, or
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunisation, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
SQ Sequence 3930 BP; 890 A; 1365 C; 1214 G; 461 T; 0 U; 0 Other;

Query Match 99.2%; Score 2436.4; DB 10; Length 3930;
Best Local Similarity 100.0%; Pred. No. 3.6e-293;
Matches 2437; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 14 TGGCCGAGGCGCATGAGCCAGGCGCCAGGCGCCAGTCATGAGCGGCGCACTTCA 73
DB 1487 TCGCCGAGGCGCATGAGCGCCAGGCGCCAGGCGCCAGTCATGAGCGGCGCACTTCA 1546
QY 74 AGGGCCCCAAGGCGCATGAGTCCTTCACTGCGGCAAGGAGGCCCATGCGCCCA 133
DB 1547 AGGGCCCCAAGGCGCATGAGTCCTTCACTGCGGCAAGGAGGCCCATGCGCCCA 1606
QY 134 ACTGCG 193
DB 1607 ACTGCG 1666
QY 194 AGGACTGCAACGAGCG 253
DB 1667 AGGACTGCAACGAGCG 1726
QY 254 AGGCGCGCGAGTTCCCG 313
DB 1727 AGGCGCGCGAGTTCCCG 1786
QY 314 AGGTGCG 373
DB 1787 AGGTGCG 1846
QY 374 TCCCGCGAGTCACTCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
DB 1847 TCCCGCGAGTCACTCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906
QY 434 AGGAGCGCGCTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 493
DB 1907 AGGAGCGCGCTGTGAGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
QY 494 GCAAGTGAAGCG 553
DB 1967 GCAAGTGAAGCG 2026
QY 554 ACCAGATCTGATGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 613
DB 2027 ACCAGATCTGATGAGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2086
QY 614 CCCCCTGGAACATGCG 673

DB 2087 CCCCCTGGAACATGCG 2146
QY 674 CCATCAAGCGCGCATGAGACCGTGGCCCGTGAAGCTGAAGCCCGGCGCATGAGACCGCGCGCAAGG 733
DB 2147 CCATCAAGCGCGCATGAGACCGTGGCCCGTGAAGCTGAAGCCCGGCGCATGAGACCGCGCGCAAGG 2206
QY 734 TGAAGCAGTGGCG 793
DB 2207 TGAAGCAGTGGCG 2266
QY 794 TGAAGAGAGGCG 853
DB 2267 TGAAGAGAGGCG 2326
QY 854 TCGCGATCAAGAAGAAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 913
DB 2327 TCGCGATCAAGAAGAAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2386
QY 914 ACAAGCGCACCGAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGACCGCGCGCGCGCG 973
DB 2387 ACAAGCGCACCGAGGACTTCTGGGAGGTGAGCTGGGCGATCCCGACCGCGCGCGCGCG 2446
QY 974 AGAAGAGAGAGCGGTGACCGGTGCTGAGCGGTGCGAGCGCTTCAAGCGTGCCTGG 1033
DB 2447 AGAAGAGAGAGCGGTGACCGGTGCTGAGCGGTGCGAGCGCTTCAAGCGTGCCTGG 2506
QY 1034 ACGAGGACTTTCGCGCAAGTACACCGCGCTTACACATCCCGCGCATCAAGACGAGACCGCG 1093
DB 2507 ACGAGGACTTTCGCGCAAGTACACCGCGCTTACACATCCCGCGCATCAAGACGAGACCGCG 2566
QY 1094 GCATTCGCTACCAAGTACCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153
DB 2567 GCATTCGCTACCAAGTACCAAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2626
QY 1154 AGAGCAGATGACCAAGATCTTGAAGCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1213
DB 2627 AGAGCAGATGACCAAGATCTTGAAGCCCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2686
QY 1214 ACCAGGCG 1273
DB 2687 ACCAGGCG 2746
QY 1274 AGGAGCTGCGGAGGACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1333
DB 2747 AGGAGCTGCGGAGGACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2806
QY 1334 AGAGCG 1393
DB 2807 AGAGCG 2866
QY 1394 AGCTGCGCGGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGAGCTGTGGCGAAGCTGA 1453
DB 2867 AGCTGCGCGGAGAGAGAGCTGAGACCGTGAACGACATCCAGAGAGCTGTGGCGAAGCTGA 2926
QY 1454 ACTGGGCGAGCGAGATCAACCGCGCGCATCAAGGTGCGCGAGCTGTGCGAGCGCG 1513
DB 2927 ACTGGGCGAGCGAGATCAACCGCGCGCATCAAGGTGCGCGAGCTGTGCGAGCGCG 2986
QY 1514 GCGCGCAAGCGCTGACCGAGCATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1573
DB 2987 GCGCGCAAGCGCTGACCGAGCATGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3046
QY 1574 AGAAGCGGAGATCTGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1633
DB 3047 AGAAGCGGAGATCTGTGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3106
QY 1634 TGTGCGCGGAGATCCAGAGAGAGGCGCGAGCGAGTGAACCTTACAGAGATCTTACAGAGAGC 1693
DB 3107 TGTGCGCGGAGATCCAGAGAGAGGCGCGAGCGAGTGAACCTTACAGAGATCTTACAGAGAGC 3166
QY 1694 CTTTCAAGACCTGAGAGACCGCGCAAGTACCGCGCAAGATGCGGCGCGCGCGCGCGCGCG 1753
DB 3167 CTTTCAAGACCTGAGAGACCGCGCAAGTACCGCGCAAGATGCGGCGCGCGCGCGCGCGCG 3226

601 CTGATGGCCCCCCTGGTAAATCATTCGGCCGCAACATGTGACCCAGCTGGCTGC 660
601 CTGATGGCCCCCCTGGTAAATCATTCGGCCGCAACATTCGTGACCCAGCTGGCTGC 660
661 ACCCTGAATCTTCCCATGAGCCCATTCGAGACCGTGGCCCGTGAACCTGAAGCCCGGCTG 720
661 ACCCTGAATCTTCCCATGAGCCCATTCGAGACCGTGGCCCGTGAACCTGAAGCCCGGCTG 720
721 GACGGCCCCAAGGTGAAGCATGTGGCCCTGACCCGAGAGAAAGATCAAGGCCCTGACCGCC 780
721 GACGGCCCCAAGGTGAAGCATGTGGCCCTGACCCGAGAGAAAGATCAAGGCCCTGACCGCC 780
781 ATCTCGAGAGATGAGAGAGAGGCAAGATCAACCAAGATCGCCCTCGAGAACCCCTTAC 840
781 ATCTCGAGAGATGAGAGAGAGGCAAGATCAACCAAGATCGCCCTCGAGAACCCCTTAC 840
841 AACACCCCTGTGTTCCTCATCAAGAGAGAGACACCAAGTGGCCGCAAGCTGTGTGAC 900
841 AACACCCCTGTGTTCCTCATCAAGAGAGAGACACCAAGTGGCCGCAAGCTGTGTGAC 900
901 TTCCGAGAGCTGAACAGAGGCAACCAAGGCTTCTGGAGGTGCACTGGGCAATCCCGAC 960
901 TTCCGAGAGCTGAACAGAGGCAACCAAGGCTTCTGGAGGTGCACTGGGCAATCCCGAC 960
961 CCCGCGGCTTGAAGAGAGAGGCGTGAACCGTGTGAGAGTGGGCGAACCGCTTACTTC 1020
961 CCCGCGGCTTGAAGAGAGAGGCGTGAACCGTGTGAGAGTGGGCGAACCGCTTACTTC 1020
1021 AGCGTGGCCCTGGAGAGAGAGCTTCCGCAAGTCAACCGGCTTCAACATCCCGACATCAAC 1080
1021 AGCGTGGCCCTGGAGAGAGAGCTTCCGCAAGTCAACCGGCTTCAACATCCCGACATCAAC 1080
1021 AGCGTGGCCCTGGAGAGAGAGCTTCCGCAAGTCAACCGGCTTCAACATCCCGACATCAAC 1080
1081 AACGAGACCCCGGATCCGCTACCAAGTCAACAGTGTGACCCGAGGCTGAGAGGCGACG 1140
1081 AACGAGACCCCGGATCCGCTACCAAGTCAACAGTGTGACCCGAGGCTGAGAGGCGACG 1140
1141 CCCAGCATCTTCCAGAGAGAGATGAACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCC 1200
1141 CCCAGCATCTTCCAGAGAGAGATGAACCAAGATCTGAGAGCCCTTCCGCGCCGCAACCC 1200
1201 GAGATCGATCTTCAAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1260
1201 GAGATCGATCTTCAAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1260
1201 GAGATCGATCTTCAAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1260
1261 CGCGCAGAGATCGAGAGAGCTGGCAAGCACTGTGCGTGGGCTTCAACACCCCGAC 1320
1261 CGCGCAGAGATCGAGAGAGCTGGCAAGCACTGTGCGTGGGCTTCAACACCCCGAC 1320
1321 AAGAGACACAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1374
1321 AAGAGACACAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1374
1321 AAGAGACACAGAGAGAGCCCTTCTGATCGTGGCAACGACCTGGAGATTCGGACAC 1380
1375 TGGACCGTGGCAAGCCATGAGAGCTGGCCGAGAGAGAGAGCTGGAACGATTCGAC 1434
1375 TGGACCGTGGCAAGCCATGAGAGCTGGCCGAGAGAGAGAGCTGGAACGATTCGAC 1434
1381 TGGACCGTGGCAAGCCATGAGAGCTGGCCGAGAGAGAGAGCTGGAACGATTCGAC 1440
1435 AAGCTGTGGGCAAGCTGAACCTGGGCAAGCCAGATCTACCCCGGATCAAGGTGGCCAG 1494
1435 AAGCTGTGGGCAAGCTGAACCTGGGCAAGCCAGATCTACCCCGGATCAAGGTGGCCAG 1494
1441 AAGCTGTGGGCAAGCTGAACCTGGGCAAGCCAGATCTACCCCGGATCAAGGTGGCCAG 1500
1495 CTGTGCAAGCTGTGCTGGCGGCGCAAGGCTTGAACGATCTGTGCTGTGCTGTGACGAGAG 1554
1501 CTGTGCAAGCTGTGCTGGCGGCGCAAGGCTTGAACGATCTGTGCTGTGCTGTGACGAGAG 1560
1555 GCGGAGCTGGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGTGTGACGCGCTGAC 1614
1555 GCGGAGCTGGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGTGTGACGCGCTGAC 1614
1561 GCGGAGCTGGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGTGTGACGCGCTGAC 1620
1561 GCGGAGCTGGAGAGCTGGCGGAGAACCGCGAGATCTGTGCGGAGCCGTGTGACGCGCTGAC 1620
1615 TACGACCCCGAGAGAGAGCTGTGGCCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
1615 TACGACCCCGAGAGAGAGCTGTGGCCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
1621 TACGACCCCGAGAGAGAGCTGTGGCCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
1621 TACGACCCCGAGAGAGAGCTGTGGCCGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
1675 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1734
1675 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1734

1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
1735 ACCGCGCAACCAAG 1794
1741 ACCGCGCAACCAAG 1800
1795 AGCATGTGATCTGGGCGAG 1854
1801 AGCATGTGATCTGGGCGAG 1860
1855 GAGACCTGTGTGAG 1914
1861 GAGACCTGTGTGAG 1920
1915 ACCGCGCGGCTGTGTGAG 1974
1921 ACCGCGCGGCTGTGTGAG 1980
1975 ACCTTCTACGTGTGAG 2034
1981 ACCTTCTACGTGTGAG 2040
2035 ACCGAGCGGCGCGGAG 2094
2041 ACCGAGCGGCGCGGAG 2100
2095 CTGAG 2154
2101 CTGAG 2160
2155 AGCGAGTACGCGCTGTGTGAG 2214
2161 AGCGAGTACGCGCTGTGTGAG 2220
2215 AACGAGATCATGAG 2274
2221 AACGAGATCATGAG 2280
2275 CACAG 2334
2281 CACAG 2340
2335 GTGCTGTCTGTGAG 2394
2341 GTGCTGTCTGTGAG 2400
2395 TACGTGGGCAAGCGGCGCTTGAAGATGATTAAAGCTTCCGCGGCTTGAAGAGAGAGAGAG 2454
2401 TACGTGGGCAAGCGGCGCTTGAAGATGATTAAAGCTTCCGCGGCTTGAAGAGAGAGAGAG 2460
2455 TTC 2457
2461 TTC 2463

RESULT 7
ADM73765
ID ADM73765 standard; DNA; 2463 BP.
XX
XX ADM73765;
XX
XX 03-JUN-2004 (first entry)
XX
XX
XX HIV-1 polynucleotide #8.
XX
XX HIV-1; gene; ds; HIV pol; immune response; DNA immunization;
XX
XX HIV type C protein; immunostimulant.
XX
XX Human immunodeficiency virus 1.
XX
XX US200323961-A1.

D	04-DEC-2003.
XX	
P	05-JUL-2001; 2001US-00899575.
XX	
P	05-JUL-2000; 2000US-00610313.
XX	
P	(MEGE/) MEGE J Z.
XX	
PA	(BARN/) BARNETT S W.
PA	(ENGE/) ENGELBRECHT S.
PA	(RENS/) RENSBUERG E J V.
XX	
PI	Megede JZ, Barnett SW, Engelbrecht S, Rensburg EJV;
DR	WPI; 2004-060515/06.
PT	New expression cassette comprising a polynucleotide sequence encoding an
PT	HIV Pol polypeptide, useful in eliciting an immune response, in DNA
PT	immunization, generating of packaging cell lines or in producing HIV Type
FT	C proteins.
XX	
PS	Claim 1; SEQ ID NO 31; 160pp; English.
CC	The invention relates to an expression cassette comprising a
CC	polynucleotide sequence encoding an HIV Pol polypeptide. The invention
CC	also relates to a recombinant expression system for use in a host cell
CC	comprising an expression cassette, where the polynucleotide sequence
CC	further comprises control elements capable of driving expression in the
CC	selected host cell, a cell comprising an expression cassette where the
CC	polynucleotide sequence further comprises control elements compatible
CC	with the expression in the cell and a composition for generating an
CC	immunological response, comprising an expression cassette. The expression
CC	cassette and the method of the invention are useful in eliciting an
CC	immune response, in DNA immunisation, in generation of packaging cell
CC	lines and in producing HIV Type C proteins. This sequence represents an
CC	HIV-1 polynucleotide of the invention.
XX	
SQ	Sequence 2463 BP; 567 A; 835 G; 759 G; 302 T; 0 U; 0 Other;
Query Match	99.2%; Score 2436.2; DB 12; Length 2463;
Best Local Similarity	99.6%; Pred. No. 4.1e-293;
Matches 2454; Conservative	0; Mismatches 3; Indels 6; Gaps 1
QY	1 GTGACGGCACCATTGAGCGGCGCATGAGCCAGGCCACGAGGCCCAATCTGTATGCAG 60
DB	1 GTGACGGCACCATGAGCGGCGCATGAGCCAGGCCACGAGGCCCAATCTGTATGCAG 60
QY	61 CGCAGCAACTTCAGAGGGCCCCAAGGCGCATCATCATAGTGCTTAACTGCGGCAAGAAGGCG 120
DB	61 CGCAGCAACTTCAGAGGGCCCCAAGGCGCATCATCATAGTGCTTAACTGCGGCAAGAAGGCG 120
QY	121 CACATGCGCCGCAATGCGCGGCGCCCCCGCAAGGAAGGGGTGTGTGAAGTGCAGCAAGAG 180
DB	121 CACATGCGCCGCAATGCGCGGCGCCCCCGCAAGGAAGGGGTGTGTGAAGTGCAGCAAGAG 180
QY	181 GGCCACACAGATGAGAGACTGACCGAGAGCGCCAGGCCAATTTCTTCGCGAGGAACTTGCGC 240
DB	181 GGCCACACAGATGAGAGACTGACCGAGAGCGCCAGGCCAATTTCTTCGCGAGGAACTTGCGC 240
QY	241 TTCCCCCAGGGGCAAGGCCCCGCGAGTTCCCAGCGAGACAGAACTCGGCGCAAGCCCCCACC 300
DB	241 TTCCCCCAGGGGCAAGGCCCCGCGAGTTCCCAGCGAGACAGAACTCGGCGCAAGCCCCCACC 300
QY	301 AGCGCGAGAGCTGACAGGTGCGGGCGAACACCCCGAGCGAGAGGCGCGGCGGCGGCGGCG 360
DB	301 AGCGCGAGAGCTGACAGGTGCGGGCGAACACCCCGAGCGAGAGGCGCGGCGGCGGCGGCGG 360
QY	361 GGCACCTGGAATTTCCCCCAATACCTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420
DB	361 GGCACCTGGAATTTCCCCCAATACCTCTGTGGCAGCGCCCCCTGTGTGAGCATCAAGGTG 420
QY	421 GCGCGGCAGATCAAGAGGCGCTGTGTGACAACCGGCGCGCATGACACCTGTGTGAGAGAG 480
DB	421 GCGCGGCAGATCAAGAGGCGCTGTGTGACAACCGGCGCGCATGACACCTGTGTGAGAGAG 480

QY	481	ATGAGCTGCCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGGCTTATCAAG	540
Db	481	ATGAGCCCTGCCCCGGCAAGTGGAAAGCCCAAGATGATCGGCGGCATCGCGGCTTATCAAG	540
QY	541	GTGGCGCAGTACGACCAAGATCTTAATGAGATCTGCGGCAAGAGGCGATCGGCACGTG	600
Db	541	GTGGCGCAGTACGACCAAGATCTTAATGAGATCTGCGGCAAGAGGCGATCGGCACGTG	600
QY	601	CTGATCGGCCCCCAACCCCGGTGAAATCATATCGGCGGCAATGCTGACCCAGCTGGGCTGC	660
Db	601	CTGATCGGCCCCCAACCCCGGTGAAATCATATCGGCGGCAATGCTGACCCAGCTGGGCTGC	660
QY	661	ACCTGAACTTCCCCATCAGGCCCATGAGACCGTGCCCGGTGAAGCTGAAGCCCGGATG	720
Db	661	ACCTGAACTTCCCCATCAGGCCCATGAGACCGTGCCCGGTGAAGCTGAAGCCCGGATG	720
QY	721	GACGCGCCCAAGTGGAAAGAGTGGCCCCCTGACCCAGAGGAAGATCAAGGCCCTGACCGC	780
Db	721	GACGCGCCCAAGTGGAAAGAGTGGCCCCCTGACCCAGAGGAAGATCAAGGCCCTGACCGC	780
QY	781	ATCTGCGAGAGATGAGAGAGAGAGGCGCAATTCACAAAGATCGGCCCCGAGAACCCCTTAC	840
Db	781	ATCTGCGAGAGATGAGAGAGAGAGGCGCAATTCACAAAGATCGGCCCCGAGAACCCCTTAC	840
QY	841	AAACACCCCGGATTCGCGCATCAAGAAAGAAAGACAGCAACAAAGTGGCGCAAGCTGTGGAC	900
Db	841	AAACACCCCGGATTCGCGCATCAAGAAAGAAAGACAGCAACAAAGTGGCGCAAGCTGTGGAC	900
QY	901	TTCCGCGAGCTGAACAAAGCGCACCCAGAGCTTCTTGGAGAGTGCACTTGGGATCCCCAC	960
Db	901	TTCCGCGAGCTGAACAAAGCGCACCCAGAGCTTCTTGGAGAGTGCACTTGGGATCCCCAC	960
QY	961	CCCCCGGCGCTGAAGAAAGAAAGAGGCTGACCGTGCTGACCGTGGGCGAAGCGCTTACTT	1020
Db	961	CCCCCGGCGCTGAAGAAAGAAAGAGGCTGACCGTGCTGACCGTGGGCGAAGCGCTTACTT	1020
QY	1021	AGCGTGCCCTCGAGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGATCAAC	1080
Db	1021	AGCGTGCCCTCGAGACGAGGACTTCCGCAAGTACACCGCTTCAACATCCCGAGATCAAC	1080
QY	1081	AAACGAGACCCCGGCGATCGCTTACCAAGTCAACGTGCTGCGCCAGAGGCTTGAAGGGCAGC	1140
Db	1081	AAACGAGACCCCGGCGATCGCTTACCAAGTCAACGTGCTGCGCCAGAGGCTTGAAGGGCAGC	1140
QY	1141	CCAGGACATCTTCCAGAGGAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCC	1200
Db	1141	CCAGGACATCTTCCAGAGGAGCATGACCAAGATCTTGGAGCCCTTCCGCGCCCGCAACCC	1200
QY	1201	GAGATCGTGAATCTACAGGCCCCCTCTTACGTGGGCAAGGACCTTGAAGATCGGCCAGAC	1260
Db	1201	GAGATCGTGAATCTACAGGCCCCCTCTTACGTGGGCAAGGACCTTGAAGATCGGCCAGAC	1260
QY	1261	CGCGCGAAGATCGAGAGAGCTGGGCAAGACCTTGCTGCGCTGGGGCTTACCAACCCCGAC	1320
Db	1261	CGCGCGAAGATCGAGAGAGCTGGGCAAGACCTTGCTGCGCTGGGGCTTACCAACCCCGAC	1320
QY	1321	AAGAAGACCAAGAAAGAGCCCCCTCTTCTGCGCAT-----CGAGCTGCAACCCGACAG	1374
Db	1321	AAGAAGACCAAGAAAGAGCCCCCTCTTCTGCGCAT-----CGAGCTGCAACCCGACAG	1380
QY	1375	TGGAACCGTGCAGCCCATGAGCTGCGCGAAGAAAGAGCTGGAACCTGTAAACGACATCCAG	1433
Db	1381	TGGAACCGTGCAGCCCATGAGCTGCGCGAAGAAAGAGCTGGAACCTGTAAACGACATCCAG	1440
QY	1435	AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTAACCCGGCATCAAGTGTGCGCAG	1494
Db	1441	AAGCTGTGGGCAAGCTGAATCTGGGCGCAGCAGATCTAACCCGGCATCAAGTGTGCGCAG	1500
QY	1495	CTGTGCAAGCTGTGCGCGCGCGCAAGGCCCTTGACCGACATGTGTGCTTGAACGAGAG	1554
Db	1501	CTGTGCAAGCTGTGCGCGCGCGCGCAAGGCCCTTGACCGACATGTGTGCTTGAACGAGAG	1560

```

Oy 1555 GCCGAGCTGAGCTGAGCGAGAACCGCGAGATCTCTGCGAGAGCCCGTCACAGCGCTGTAC 1614
Db 1561 GCCGAGCTGAGCTGAGCGAGAACCGCGAGATCTCTGCGAGAGCCCGTCACAGCGCTGTAC 1620
Oy 1615 TACGACCCCGAGAGAGAGCTGTGTGCGCGAGATCCGAGAGCGAGCGACGACAGTGGACC 1674
Db 1621 TACGACCCCGAGAGAGAGCTGTGTGCGCGAGATCCGAGAGCGAGCGACGACAGTGGACC 1680
Oy 1675 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAGAGACCGGCAAGTAAGCCAGATGGGC 1734
Db 1681 TACGAGATCTACAGAGAGCCCTTCAAGAACCTGAGAGACCGGCAAGTAAGCCAGATGGGC 1740
Oy 1735 ACCGCCCAACCAACGAGCTGAGAGAGCTGACCGAGCGCTGTGACAGAGATCGCCATGGAG 1794
Db 1741 ACCGCCCAACCAACGAGCTGAGAGAGCTGACCGAGCGCTGTGACAGAGATCGCCATGGAG 1800
Oy 1795 AGCATCTGTGATCTGGGGGAGAGACCCCGAAGTTCCCGCTGCGCCATCCAGAGAGAGCTGG 1854
Db 1801 AGCATCTGTGATCTGGGGGAGAGACCCCGAAGTTCCCGCTGCGCCATCCAGAGAGAGCTGG 1860
Oy 1855 GAGACCTGTGAGACCGACTAGTGGAGGCGACCTGATCCCGAGTGGGAGTTGTGTAAAC 1914
Db 1861 GAGACCTGTGAGACCGACTAGTGGAGGCGACCTGATCCCGAGTGGGAGTTGTGTAAAC 1920
Oy 1915 ACCCCCCCTGTGTGAGAGCTGTGTACCAAGCTGAGAGAGAGCCCATCATCGGCGCCGAG 1974
Db 1921 ACCCCCCCTGTGTGAGAGCTGTGTGTACCAAGCTGAGAGAGAGCCCATCATCGGCGCCGAG 1980
Oy 1975 ACCTTCTACGTGAGAGCGGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTTG 2034
Db 1981 ACCTTCTACGTGAGAGCGGCGCGCCCAACCGGAGACCAAGATCGGCAAGCGCGCTACGTTG 2040
Oy 2035 ACCGACCGGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAG 2094
Db 2041 ACCGACCGGGGCGCGGAGAGATCGTGAAGCTGACCGAGACCAACCAAGAGAGCGAG 2100
Oy 2095 CTGCAAGGCATCTCAAGCTGCGCTGTGAGAGACAGCGGCAAGAGTGAACATCGTGAACGAC 2154
Db 2101 CTGCAAGGCATCTCAAGCTGCGCTGTGAGAGACAGCGGCAAGAGTGAACATCGTGAACGAC 2160
Oy 2155 AGCCAGTACGCTTGGGCGATCATCCAGGCGCCAGCCCGACAGAGCGAGAGCGAGCTGTGTG 2214
Db 2161 AGCCAGTACGCTTGGGCGATCATCCAGGCGCCAGCCCGACAGAGCGAGAGCGAGCTGTGTG 2220
Oy 2215 AACGAGATCATGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCC 2274
Db 2221 AACGAGATCATGAGCAGCTGATCAAGAGAGAGAGTGTACCTGAGCTGGGTGCGCGCC 2280
Oy 2275 CACAAAGGCGATCGCGGCAACGAGCAGATCGAACAGCTGTGTGAGAGAGGCGCATTCGCAAG 2334
Db 2281 CACAAAGGCGATCGCGGCAACGAGCAGATCGAACAGCTGTGTGAGAGAGGCGCATTCGCAAG 2340
Oy 2335 GTGCTGTCTCTGGAAGGCGATCGATGGCGGCGATCTGTATCTACAGTACATGAGCACTTG 2394
Db 2341 GTGCTGTCTCTGGAAGGCGATCGATGGCGGCGATCTGTATCTACAGTACATGAGCACTTG 2400
Oy 2395 TACGTGGGAGAGCGGCGCGCTTGAATCGATTAAAGCTTCCGCGGCGTAAAGCAGCGGTAA 2454
Db 2401 TACGTGGGAGAGCGGCGCGCTTGAATCGATTAAAGCTTCCGCGGCGTAAAGCAGCGGTAA 2460
Oy 2455 TTC 2457
Db 2461 TTC 2463

```

RESULT 8
ADCI3231
ID ADCI3231 standard; DNA; 3930 BP.

XX AC ADCI3231;
XX DT 18-DEC-2003 (first entry)
XX

```

DE DNA of HIV construct GagComp1PolmutatC_SHQ ID NO 10.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RaseH; Pol; Tat; ProC;
KM Rev; HIV type C; gene therapy; vaccine; immunization; HIV; ds.
XX
XX Human immunodeficiency virus.
XX
XX WO200304620-A2.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002WO-US021420.
XX
XX 05-JUL-2001; 2001US-0303192P.
XX
XX 31-AUG-2001; 2001US-0316860P.
XX
XX 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (UYST-) UNIV STIELENBOSCH.
XX
XX Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg BJ,
XX
XX WPI; 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RaseH, Pol, Tat,
XX
XX ProC, or Rev polypeptide, useful for immunization, or generating
XX
XX packaging cell lines.
XX
XX Disclosure; Fig 7; 301pp; English.
XX
XX
XX The invention relates to a novel expression cassette comprising a
XX
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX
XX Int, Nef, p15RaseH, Pol, Tat, ProC, or Rev polypeptide. The novel
XX
XX expression cassette can be used to treat HIV type C by gene therapy or
XX
XX used in the development of a vaccine. The gene delivery vector is
XX
XX administered intramuscularly, intravenously, intranasally,
XX
XX subcutaneously, intradermally, transdermally, intravaginally,
XX
XX intrarectally, orally or intravenously. The expression cassette is useful
XX
XX for immunisation, generating packaging cell lines and producing HIV
XX
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX
XX Type C related sequence of the invention.
XX
XX
XX Sequence 3930 BP; 889 A; 1365 C; 1214 G; 462 T; 0 U; 0 Other;
XX
XX
XX Query Match 99.1%; Score 2434.8; DB 10; Length 3930;
XX
XX Best Local Similarity 99.9%; Fred. No. 5.7e-293;
XX
XX Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 14 TGCCGAGGCGCATGAGCGAGCGACAGCGCGCAACATCTGTATGAGCGAGCAACTTCA 73
Db 1487 TGCCGAGGCGCATGAGCGAGCGACAGCGCGCAACATCTGTATGAGCGAGCAACTTCA 1546
Oy 74 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATCGCGGCAAGAGAGCGCCATCGCCGCA 133
Db 1547 AGGCGCCCAAGCGCATCATCAAGTGTCTTCAATCGCGGCAAGAGAGCGCCATCGCCGCA 1606
Oy 134 ACTGCGCGCGCCCGCGCAAGAGAGGCTGTGTGAAGTGTGCGCAAGAGAGCGCCACAGATGA 193
Db 1607 ACTGCGCGCGCCCGCGCAAGAGAGGCTGTGTGAAGTGTGCGCAAGAGAGCGCCACAGATGA 1666
Oy 194 AGGACTGACCGAGCGCGCGCGCAACTTCTTCCGCGAGAGACTGGCTTCCCGCAGGCA 253
Db 1667 AGGACTGACCGAGCGCGCGCGCAACTTCTTCCGCGAGAGACTGGCTTCCCGCAGGCA 1726
Oy 254 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCAACAGCGCCACAGCGCGAGAGTGC 313
Db 1727 AGGCGCGGAGTTCCCGAGCGAGAGAACCGCGCAACAGCGCCACAGCGCGAGAGTGC 1786
Oy 314 AGGTGCGCGGAGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCACT 373
Db 1787 AGGTGCGCGGAGCAACCCCGCAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCACT 1846

```


XX 18-DEC-2003 (first entry)
DT DNA of HIV construct GagCompIpolmImltina_C SEQ ID NO 11.
DE
XX expression cassette; HIV Gag; Env; Int; Nef; p15KaseH; Pol; Tat; Prot;
KM Rev; HIV type C; gene therapy; vaccine; immunisation; HIV; ds.
XX Human immunodeficiency virus.
OS
XX MO2003004620-A2.
XX
XX 16-JAN-2003.
PD
XX 05-JUL-2002; 2002MO-US021420.
PF
XX 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR) CHIRON CORP.
PA (UNST-) UNIV STELLENBOSCH.
XX
PI Zur Megede J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ,
XX
DR WPI; 2003-221593/21.
XX
XX
PT New expression cassette comprising a polynucleotide sequence encoding a
PT polypeptide including an HIV Gag, Env, Int, Nef, p15KaseH, Pol, Tat,
PT Prot, or Rev polypeptide, useful for immunization, or generating
PT packaging cell lines.
XX
XX Diecloare; Fig 8; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
CC polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
CC Int, Nef, p15KaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
CC expression cassette can be used to treat HIV type C by gene therapy or
CC used in the development of a vaccine. The gene delivery vector is
CC administered intramuscularly, intravenously, intranasally,
CC subcutaneously, intradermally, transdermally, intravaginally,
CC intrarectally, orally or intravenously. The expression cassette is useful
CC for immunisation, generating packaging cell lines and producing HIV
CC polypeptides. This polynucleotide sequence represents the DNA of an HIV
CC Type C related sequence of the invention.
XX
SQ Sequence 3930 BP; 889 A; 1366 C; 1214 G; 461 T; 0 U; 0 Other;
Query Match 99.1%; Score 2434.8; DB 10; Length 3930;
Best Local Similarity 99.9%; Pred. No. 5.7e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 1787 AGGTGCGGCGGCAACACCCCGCAGCGAGCGGCGCGGACGCGCAAGGCAACCTGAACT 1846
Qy 374 TCCCCGAGATACCTCTGTGGAGGCGCCCTCTGTGAGATATGAAGTGGCGGCAAGATCA 433
Db 1847 TCCCCGAGATACCTCTGTGGAGGCGCCCTCTGTGAGATATGAAGTGGCGGCAAGATCA 1906
Qy 434 AGGAGGCGCTGTGGACACCGGCGCGGACGACACCTGTCTGAGAGATGAGACCTGCGCG 493
Db 1907 AGGAGGCGCTGTGGACACCGGCGCGGACGACACCTGTCTGAGAGATGAGACCTGCGCG 1966
Qy 494 GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTATCAAGTGGCGCAAGTACG 553
Db 1967 GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTATCAAGTGGCGCAAGTACG 2026
Qy 554 ACCAGATCTGATGAGATCTGCGGCAAGAGGCAATGTGGACCGTGTGATTCGGCCCCA 613
Db 2027 ACCAGATCTGATGAGATCTGCGGCAAGAGGCAATGTGGACCGTGTGATTCGGCCCCA 2086
Qy 614 CCCCCGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGCACCTGAACTTCC 673
Db 2087 CCCCCGTGAACATCATCGGCGGCAACATGTGACCCAGCTGGGCTGCACCTGAACTTCC 2146
Qy 674 CCATCAAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGAGCGGCGCAAGG 733
Db 2147 CCATCAAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCCCGGCAATGAGCGGCGCAAGG 2206
Qy 734 TGAAGCAGTGGCCCTCTGACCGAGAGAAATCAAGGCTCTGACCGCATCTGCGAGAGGA 793
Db 2207 TGAAGCAGTGGCCCTCTGACCGAGAGAAATCAAGGCTCTGACCGCATCTGCGAGAGGA 2266
Qy 794 TGAAGAGAGAGGAGATCAACCAAGATCGGCGCGGCAAGACCCCTGACACACCCCGGTGT 853
Db 2267 TGAAGAGAGAGGAGATCAACCAAGATCGGCGCGGCAAGACCCCTGACACACCCCGGTGT 2326
Qy 854 TCGCCATCAAGAGAGAGAGACACCAAGTGGCGGCAAGCTGTGATCTTCGCGAGCTGA 913
Db 2327 TCGCCATCAAGAGAGAGAGACACCAAGTGGCGGCAAGCTGTGATCTTCGCGAGCTGA 2386
Qy 914 ACAAGCGGACCCAGAGACTTCTGGAGAGTGGAGCTGGGCAATCCCCACCCCGCGGCTGA 973
Db 2387 ACAAGCGGACCCAGAGACTTCTGGAGAGTGGAGCTGGGCAATCCCCACCCCGCGGCTGA 2446
Qy 974 AGAAGAGAGAGAGGCTGACGCTGTGAGCTGTGAGCGTGGGAGACGCTTACAGCTGACCTGTGG 1033
Db 2447 AGAAGAGAGAGAGGCTGACGCTGTGAGCTGTGAGCGTGGGAGACGCTTACAGCTGACCTGTGG 2506
Qy 1034 ACAGAGCTTCGCAAGTACACCGGCTTCAACATTCGCCAGATCAACAGAGACCCCGG 1093
Db 2507 ACAGAGCTTCGCAAGTACACCGGCTTCAACATTCGCCAGATCAACAGAGACCCCGG 2566
Qy 1094 GCATTCGCTTACAGTACACCGTGTGCGGCGGAGGCTGTGAAGGCGAGCCCGCAGATCTTCC 1153
Db 2567 GCATTCGCTTACAGTACACCGTGTGCGGCGGAGGCTGTGAAGGCGAGCCCGCAGATCTTCC 2626
Qy 1154 AGAGCAGATACCAAGATCTGTGAGCCCTTCCGCGCGCGGCAACCCCGAGATCTGTATCT 1213
Db 2627 AGAGCAGATACCAAGATCTGTGAGCCCTTCCGCGCGCGGCAACCCCGAGATCTGTATCT 2686
Qy 1214 ACAGGCGCCCTGTGACGTGGGCAAGGACCTGAGATCGGCAAGACCGGCGCAAGATCG 1273
Db 2687 ACAGGCGCCCTGTGACGTGGGCAAGGACCTGAGATCGGCAAGACCGGCGCAAGATCG 2746
Qy 1274 AGAGCTGCGGAGACACTGTGTGGCTGGGGCTTACACACCCCGGCAAGAGAGACCGGA 1333
Db 2747 AGAGCTGCGGAGACACTGTGTGGCTGGGGCTTACACACCCCGGCAAGAGAGACCGGA 2806
Qy 1334 AGGAGCGCCCTCTCTGCGCATCGAGCTGACCCCGCAAGTGAACGCTGACGCGCATCG 1393
Db 2807 AGGAGCGCCCTCTCTGCGCATCGAGCTGACCCCGCAAGTGAACGCTGACGCGCATCG 2866
Qy 1394 AGCTGCCGAGAGAGAGAGCTGAGCGTGAACGATTCAGAAAGCTGTGTGGGCAAGCTGA 1453

Db	2867	AGCTGCCCCGAGAAAGAGACGTGAGACCGTGAACCAACATTCAGAAAGCTGTGTGGCAAGCTGA	2926
OY	1454	ACTGAGGCACGACAGATCTTACCCCGGCATCAAGGTGTGCGCAGCTGTGACAGCTGTGCGCG	1513
Db	2927	ACTGAGGCACGACAGATCTTACCCCGGCATCAAGGTGTGCGCAGCTGTGACAGCTGTGCGCG	2986
OY	1514	GGCGCAAGGCCCTGTACCGACATTCGTGCCCCCTGACCCGAGAGAGGCCGACCTGTGAAGCTTGCGCG	1573
Db	2987	GGCGCAAGGCCCTGTACCGACATTCGTGCCCCCTGACCCGAGAGAGGCCGACCTGTGAAGCTTGCGCG	3046
OY	1574	AGAACCGGAGATCCTTGCGCGAGCCCGGTGCACGGCGGTGTACTACGACCCCGACAGAGACC	1633
Db	3047	AGAACCGGAGATCCTTGCGCGAGCCCGGTGCACGGCGGTGTACTACGACCCCGACAGAGACC	3106
OY	1634	TGTTGTGCGCGAGATCCAGAGAGAGGCGCACGACAGTGGACCTTACAGATCTTACCGAGAGAC	1693
Db	3107	TGTTGTGCGCGAGATCCAGAGAGAGGCGCACGACAGTGGACCTTACAGATCTTACCGAGAGAC	3166
OY	1694	CCTTCAGAACCTTGAGAGCCGGCAAGTACGCCAAGATGTGCGCACCGGCCACACCAACGACG	1753
Db	3167	CCTTCAGAACCTTGAGAGCCGGCAAGTACGCCAAGATGTGCGCACCGGCCACACCAACGACG	3226
OY	1754	TGAAGAGGCTACCGAGGGCCGTGCAAGAGATTCGCGCATGTGAGAGAGATGTGTATCTGAGGAGCA	1813
Db	3227	TGAAGAGGCTACCGAGGGCCGTGCAAGAGATTCGCGCATGTGAGAGAGATGTGTATCTGAGGAGCA	3286
OY	1814	AGACCCCCCAATTCCGCTGTGCCATTCAGAAAGAGACCTGTGGAGACCTGTGTGAGACCGACT	1873
Db	3287	AGACCCCCCAATTCCGCTGTGCCATTCAGAAAGAGACCTGTGGAGACCTGTGTGAGACCGACT	3346
OY	1874	ACTGGCAGGCCACCTGGATTCGCCGAGTGGGAGTTCTGTGAACAACCCGCCCTGTGTGAAGC	1933
Db	3347	ACTGGCAGGCCACCTGGATTCGCCGAGTGGGAGTTCTGTGAACAACCCGCCCTGTGTGAAGC	3406
OY	1934	TGTGTGTACCACTGTGAGAGAGAGCCCATCATTCGCGCGCGAGACCTTTCATCSTGTGACCGCG	1993
Db	3407	TGTGTGTACCACTGTGAGAGAGAGCCCATCATTCGCGCGCGAGACCTTTCATCSTGTGACCGCG	3466
OY	1994	CCGCCCAACCGGAGAGACCAAGATTCGGCAAGTCCGGCTACGTGACCGACCCGGGGCCGGCGAGA	2053
Db	3467	CCGCCCAACCGGAGAGACCAAGATTCGGCAAGTCCGGCTACGTGACCGACCCGGGGCCGGCGAGA	3526
OY	2054	AGATCGTAGGCTTGCACCGAGACCAACCAACAGAAAGCCGAGCTGACGGCCATTCAGACTGG	2113
Db	3527	AGATCGTAGGCTTGCACCGAGACCAACCAACAGAAAGCCGAGCTGACGGCCATTCAGACTGG	3586
OY	2114	CCCTGCGAGACAGCGGCGACGCCAGGTGTGAACATCTGTACCGGACAGCCAGTACGCCCTCTGGGCA	2173
Db	3587	CCCTGCGAGACAGCGGCGACGCCAGGTGTGAACATCTGTACCGGACAGCCAGTACGCCCTCTGGGCA	3646
OY	2174	TCATCCAGGCCCCAGGCCGACAAAGACGAGAGCGACGTGTGTGAACCAAGATCTATCCGAGCAGC	2233
Db	3647	TCATCCAGGCCCCAGGCCGACAAAGACGAGAGCGACGTGTGTGTGAACCAAGATCTATCCGAGCAGC	3706
OY	2234	TGATCAAGAGGAGAGGTGTACTCTGAGCTGTGTGCCCCGACCAAGAGGCAATCCGCGGCA	2293
Db	3707	TGATCAAGAGGAGAGGTGTACTCTGAGCTGTGTGCCCCGACCAAGAGGCAATCCGCGGCA	3766
OY	2294	ACGAGCAGATTCGACAGAGCTGTGTAGCAAGGGCATTCGCAAGGTGTCTTCTGTGACGGCA	2353
Db	3767	ACGAGCAGATTCGACAGAGCTGTGTAGCAAGGGCATTCGCAAGGTGTCTTCTGTGACGGCA	3826
OY	2354	TCGATGAGGGGCAATCGTGAATCTACCGGTATCATGTGACGACCTGTGACGTGTGGGACGGCGGCC	2413
Db	3827	TCGATGAGGGGCAATCGTGAATCTACCGGTATCATGTGACGACCTGTGACGTGTGGGACGGCGGCC	3886
OY	2414	CTAGAGTCGATTTAAAGCTTCCCGGGGCTTAGCACCGGT	2451
Db	3887	CTAGAGTCGATTTAAAGCTTCCCGGGGCTTAGCACCGGT	3924

ID	ACN03591 standard; DNA; 5184 BP.
XX	
XX	ACA03591.1
XX	
D7	22-MAY-2003 (first entry)
XX	
DE	Synthetic DNA encoding immunogenic HIV peptide #74.
XX	
KW	Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KV	gene therapy; packaging cell line; humoral immune response;
XX	cellular immune response; gene delivery vector; DNA immunisation; ds.
OS	Synthetic.
XX	
PN	WO2003004657-A1.
PD	
XX	16-JAN-2003.
PF	
XX	05-JUL-2002; 2002MO-US021421.
PR	
XX	05-JUL-2001; 2001US-0303192P.
PR	31-AUG-2001; 2001US-031686OP.
PR	16-JAN-2002; 2002US-0349728P.
PR	16-JAN-2002; 2002US-0349733P.
PR	16-JAN-2002; 2002US-0349871P.
XX	
PA	(CHIR) CHIRON CORP.
PI	
XX	Zur Megede J, Barnett SW, Lian Y;
DR	
XX	WPI; 2003-221602/21.
PT	
XX	New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT	polypeptides, useful as immunogenic compositions or vaccines for
PT	generating humoral or cellular immune responses against HIV in a subject,
XX	especially humans.
PS	
XX	Example 1; Fig 79; 262pp; English.
CC	
XX	The invention describes a synthetic polynucleotide encoding 2 or more
CC	immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC	derived from different HIV subtypes. The polynucleotide is useful for
CC	immunisation, generation of packaging cell lines, or production of HIV
CC	polypeptides. The polynucleotide and its encoded proteins are useful as
CC	immunogenic compositions or vaccines for generating humoral or cellular
CC	immune responses against HIV in a subject, or for inducing neutralising
CC	antibodies against HIV. The gene delivery vector comprising the
CC	polynucleotide is also useful for DNA immunisation of, or for generating
CC	an immune response (e.g. a humoral or cellular immune response) in, a
CC	subject such as a mammal, particularly a human. This sequence encodes a
CC	human immunodeficiency virus immunogenic peptide
CC	
SO	Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;
Query Match	99.1%; Score 2434.8; DB 8; Length 5184;
Best Local Similarity	99.9%; Pred. No. 5.5e-293;
Matches 2436; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
DB	
QY	14 TGCGCAGGCGCATGAGCCAGGCCACCAGCGGCACATCTTGATGTCAGCGCAGCAACTTCA 73
Db	2741 TCGCCGAGGCATGAGCCAGGCCACCAGCGGCACATCTTGATGTCAGCGCAGCAACTTCA 28000
QY	74 AGGGCCCCGAAGCGCATCATCAAGTCCTTCAACTGCGGCAAGAAGGCGCACATCGCCGCA 133
Db	2801 AGGGCCCCGAAGCGCATCATCAAGTCCTTCAACTGCGGCAAGAAGGCGCACATCGCCGCA 28660
QY	134 ACTGCGCGGCGCCCCCGCAAGAAAGGCTGTGTAAGTGTGCGGCAAGAGGCGCACAGATGA 193
Db	2861 ACTGCGCGGCGCCCCCGCAAGAAAGGCTGTGTAAGTGTGCGGCAAGAGGCGCACAGATGA 29200
QY	194 AGGACTGACCGAAGCGCCAGAGGCGCAAATTCTTCCGCGAGAACCTTGCTTCCCGAAGGCA 253
Db	2921 AGGACTGACCGAAGCGCCAGAGGCGCAAATTCTTCCGCGAGAACCTTGCTTCCCGAAGGCA 29800

254 AGGCGCGAGTTCCCGACGAGGAGAAACCGCGCAAGGCCCAACGCGGAGCTGC 313
2981 AGGCGCGAGTTCCCGACGAGGAGAAACCGCGCAAGGCCCAACGCGGAGCTGC 3040
314 AGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCAACCTGAACT 373
3041 AGGTGCGCGGCGCAACCCCGCAGCGAGGCGCGGCGCGGAGCGCAACCTGAACT 3100
374 TCCCCAATCACTCTGTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
3101 TCCCCAATCACTCTGTGGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160
434 AGGAGGCGCGTGGGACACCGGCGCGGACGACACCGTGTGGAGGAGATGAGCGTGGCG 493
3161 AGGAGGCGCGTGGGACACCGGCGCGGACGACACCGTGTGGAGGAGATGAGCGTGGCG 3220
494 GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGCGCCAGTACG 553
3221 GCAAGTGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCAATCAAGGTGCGCCAGTACG 3280
554 ACCAATCTGTATCGAATCTTGGCGGAGAGAGGCAATGGGCAACCGTGTGATCGGCGCG 613
3281 ACCAATCTGTATCGAATCTTGGCGGAGAGAGGCAATGGGCAACCGTGTGATCGGCGCG 3340
614 CCCCCGTGAACATCATCGGCGGCAATGCTGACCGAGTGGGCTGCAACCTGAACTTCC 673
3341 CCCCCGTGAACATCATCGGCGGCAATGCTGACCGAGTGGGCTGCAACCTGAACTTCC 3400
674 CCATGAGCCCATCGAAGCCGTGCGCGTGAAGCTGAAGCCCGGCAATGGAACGCGCCAAAG 733
3401 CCATGAGCCCATCGAAGCCGTGCGCGTGAAGCTGAAGCCCGGCAATGGAACGCGCCAAAG 3460
734 TGAACGATGGGCGCTTGAACCGAGAGAGATCAAGCCCTGACCGGCAATCTTGGAGAGA 793
3461 TGAACGATGGGCGCTTGAACCGAGAGAGATCAAGCCCTGACCGGCAATCTTGGAGAGA 3520
794 TGAAGAGAGGAGGAGATCAACAGATCGGCGCGGAGAACCCCTGAACAACCCCGGAT 853
3521 TGAAGAGAGGAGGAGATCAACAGATCGGCGCGGAGAACCCCTGAACAACCCCGGAT 3580
854 TCGCATCAAG 913
3581 TCGCATCAAG 3640
914 ACAAGCGCAACCGAGAGCTTCTGGAGGTGCAAGCTGGGCAATCCCAACCCCGCGGCTGA 973
3641 ACAAGCGCAACCGAGAGCTTCTGGAGGTGCAAGCTGGGCAATCCCAACCCCGCGGCTGA 3700
974 AGAAGAGAGAGGCGTGAACCGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1033
3701 AGAAGAGAGAGGCGTGAACCGTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3760
1034 ACGAGAGCTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAGAGAGAG 1093
3761 ACGAGAGCTTCCGCAAGTACACCGGCTTCAACCATCCCGAGCATCAACAGAGAGAGAG 3820
1094 GCATCGCTACCAATGACAGCTGTGCGCGAGAGGCTTGAAGGAGAGAGAGAGAGAGAG 1153
3821 GCATCGCTACCAATGACAGCTGTGCGCGAGAGGCTTGAAGGAGAGAGAGAGAGAGAG 3880
1154 AG 1213
3881 AG 3940
1214 ACCAGAGCGCGCTGTACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1273
3941 ACCAGAGCGCGCTGTACGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4000
1274 AGGAGCTGGAG 1333
4001 AGGAGCTGGAG 4060

1334 AG 1393
4061 AG 4120
1394 AGTGGCGGAG 1453
4121 AGTGGCGGAG 4180
1454 ACTGGCGGAG 1513
4181 ACTGGCGGAG 4240
1514 GCGCAAG 1573
4241 GCGCAAG 4300
1574 AGAAGCGGAG 1633
4301 AGAAGCGGAG 4360
1634 TGGTGGCGGAGATCGAAG 1693
4361 TGGTGGCGGAGATCGAAG 4420
1694 CTTCAAG 1753
4421 CTTCAAG 4480
1754 TGAAGCAGCTGAG 1813
4481 TGAAGCAGCTGAG 4540
1814 AGAAGCGGAG 1873
4541 AGAAGCGGAG 4600
1874 ACTGGCAG 1933
4601 ACTGGCAG 4660
1934 TGTGGTACCAAGCTGAG 1993
4661 TGTGGTACCAAGCTGAG 4720
1994 CCGGCAACCGGAG 2053
4721 CCGGCAACCGGAG 4780
2054 AGATCGTGAAGCTGAG 2113
4781 AGATCGTGAAGCTGAG 4840
2114 CCTTGCAG 2173
4841 CCTTGCAG 4900
2174 TCATTCAG 2233
4901 TCATTCAG 4960
2234 TGATTCAG 2293
4961 TGATTCAG 5020
2294 ACGAGCAGATCGAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2353
5021 ACGAGCAGATCGAAGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5080
2354 TCGATGGCGGATGTGATCTTACAGATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2413
5081 TCGATGGCGGATGTGATCTTACAGATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5140
2414 CTAGAGATGATTTAAAGCTTCCGCGGCTTACACCGGT 2451

DB 5141 CTAGATCGATTAAAGCTTCCCGGAGCTAGCACCGGT 5178
RESULT 11
ADCl3279
ID ADCl3279 standard; DNA; 5184 BP.
XX ADCl3279;
XX
XX 18-DEC-2003 (first entry)
XX
XX DNA of HIV construct TatRevNefgagproline_C SEQ ID NO 58.
XX
XX expression cassette; HIV Gag; Env; Int; Nef; p15RnaseH; Pol; Tat; Prot;
XX Rev; HIV type C; gene therapy; vaccine; immunization; HIV; de.
XX
XX Human immunodeficiency virus.
XX
XX WO2003004620-A2.
XX
XX 16-JAN-2003.
XX
XX 05-JUL-2002; 2002MO-US021420.
XX
XX 05-JUL-2001; 2001US-0303192P.
XX PR 31-AUG-2001; 2001US-0316860P.
XX PR 16-JAN-2002; 2002US-0349871P.
XX
XX (CHIR) CHIRON CORP.
XX (UNST-) UNIV STELLENBOSCH.
XX
XX Zur Megele J, Barnett SW, Lian Y, Engelbrecht S, Van Rensburg EJ;
XX
XX WPI: 2003-221593/21.
XX
XX New expression cassette comprising a polynucleotide sequence encoding a
XX polypeptide including an HIV Gag, Env, Int, Nef, p15RnaseH, Pol, Tat,
XX Prot, or Rev polypeptide, useful for immunization, or generating
XX packaging cell lines.
XX
XX Disclosure; Fig 55; 301pp; English.
XX
XX The invention relates to a novel expression cassette comprising a
XX polynucleotide sequence encoding a polypeptide including an HIV Gag, Env,
XX Int, Nef, p15RnaseH, Pol, Tat, Prot, or Rev polypeptide. The novel
XX expression cassette can be used to treat HIV type C by gene therapy or
XX used in the development of a vaccine. The gene delivery vector is
XX administered intramuscularly, intravenously, intranasally,
XX subcutaneously, intradermally, transdermally, intravaginally,
XX intrarectally, orally or intravenously. The expression cassette is useful
XX for immunization, generating packaging cell lines and producing HIV
XX polypeptides. This polynucleotide sequence represents the DNA of an HIV
XX Type C related sequence of the invention.
XX
XX Sequence 5184 BP; 1139 A; 1852 C; 1610 G; 583 T; 0 U; 0 Other;
SQ
Query Match 99.1%; Score 2434.8; DB 10; Length 5184;
Best Local Similarity 99.9%; Pred. No. 5.5e-293;
Matches 2436; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 14 TGGCCGAGGCGATGAGCCAGGCGCAAGGCGCAACATCTGTATGACGCGCAGCACTTCA 73
DB 2741 TGGCCGAGGCGATGAGCCAGGCGCAAGGCGCAACATCTGTATGAGCGCGCAACTTCA 2800
QY 74 AGGCGCCGAGGCGATGAGCTTCACTGCGCGCAAGAGGCGCACTGCGCCGCA 133
DB 2801 AGGCGCCGAGGCGATGAGCTTCACTGCGCGCAAGAGGCGCACTGCGCCGCA 2860
QY 134 ACTGCGCGCGCGCGCGCAAGAGGCGCTGCTGGAAGTGGCGCAAGAGGCGCACTGATGA 193
DB 2861 ACTGCGCGCGCGCGCGCAAGAGGCGCTGCTGGAAGTGGCGCAAGAGGCGCACTGATGA 2920

QY 194 AGGACTGCAACGAGCGCGCAAGGCGCAACTCTTCCGCGAGGACCTGTGCTTCCCGCAGGCA 253
DB 2921 AGGACTGCAACGAGCGCGCAAGGCGCAACTCTTCCGCGAGGACCTGTGCTTCCCGCAGGCA 2980
QY 254 AGGCGCGCGAGTTCGCCAGGAGCAAGAACCGCGCAAGAGCGCCCAAGCGCGAGCTGC 313
DB 2981 AGGCGCGCGAGTTCGCCAGGAGCAAGAACCGCGCAAGAGCGCCCAAGCGCGAGCTGC 3040
QY 314 AGGTGCGCGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 373
DB 3041 AGGTGCGCGCGCAACCCCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3100
QY 374 TCCCCAATATACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 433
DB 3101 TCCCCAATATACCTGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3160
QY 434 AGGAGCGCGCTGTGAGCACCGCGCGCGAGCACCGCTGTGAGAGATGAGACTGTCCCG 493
DB 3161 AGGAGCGCGCTGTGCGCACCGCGCGCGAGCACCGCTGTGAGAGATGAGACTGTCCCG 3220
QY 494 GCAAGTGAAGCCCAAGATGATCGCGCGATCGCGCGCTTCAATCAAGTGTGCGCGCACTAG 553
DB 3221 GCAAGTGAAGCCCAAGATGATCGCGCGATCGCGCGCTTCAATCAAGTGTGCGCGCACTAG 3280
QY 554 ACCAATCTCTGATGAGATCTGCGCGCAAGAGCGCACTGCGCGCGCGCGCGCGCGCGCG 613
DB 3281 ACCAATCTCTGATGAGATCTGCGCGCAAGAGCGCACTGCGCGCGCGCGCGCGCGCGCG 3340
QY 614 CCCCCGTGAACATCATCGCGCGCGCAATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 673
DB 3341 CCCCCGTGAACATCATCGCGCGCGCAATGATGAGCGCGCGCGCGCGCGCGCGCGCGCG 3400
QY 674 CCATGAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCGCATGAGCGCGCGCGCAAG 733
DB 3401 CCATGAGCCCATCGAGACCGTGCCTGTGAAGCTGAAGCGCGCGCATGAGCGCGCGCGCAAG 3460
QY 734 TGAACAGTGTGCG 793
DB 3461 TGAACAGTGTGCG 3520
QY 794 TGGAGAGAGGCGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 853
DB 3521 TGGAGAGAGGCGCGCAAGATCAACAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3580
QY 854 TGGCATCAAG 913
DB 3581 TGGCATCAAG 3640
QY 914 ACAAGCGCACCGAGAGCTTCTGGAGGTGAGCTGTGCGCATCCCCACCGCGCGCGCTGA 973
DB 3641 ACAAGCGCACCGAGAGCTTCTGGAGGTGAGCTGTGCGCATCCCCACCGCGCGCGCTGA 3700
QY 974 AGAAGAGAGAGCGTGAACCGTGTGAGCGTGTGAGAGCGCTTCAAGCGTGTGCGCGCTGA 1033
DB 3701 AGAAGAGAGAGCGTGTGAACCGTGTGAGCGTGTGAGAGCGCTTCAAGCGTGTGCGCGCTGA 3760
QY 1034 ACAGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGAGCATCAACAGAGACCCCG 1093
DB 3761 ACAGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGAGCATCAACAGAGACCCCG 3820
QY 1094 GCATCGCTACCAAGTACACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1153
DB 3821 GCATCGCTACCAAGTACACCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3880
QY 1154 AGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1213
DB 3881 AGAGAGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCG 3940
QY 1214 ACCAGGCG 1273
DB 3941 ACCAGGCG 4000
QY 1274 AGAGCTGCGCAAGACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 1333

```
Db 4001 AGGAGCTGCGAGACCTGCTGCGCTGCGGCTTCAACCCCGACAAGAGCACGAGA 4060
Qy 1334 AGGAGCCCCCTTCTGCGCCATCGAGCTGCAACCCCGACAAGAGCACTGTGACCCCATCG 1393
Db 4061 AGGAGCCCCCTTCTGCGCCATCGAGCTGCAACCCCGACAAGAGCACTGTGACCCCATCG 4120
Qy 1394 AGTGCCTCCGAGAGAGAGAGCTGAGCCGTGAACGATCCAGAGAGTGTGTGGCAAGCTGA 1453
Db 4121 AGCTGCCCGAGAGAGAGAGCTGAGCCGTGAACGATCCAGAGAGTGTGTGGCAAGCTGA 4180
Qy 1454 ACTGGGCGACGCGATCTTACCCCGCATCAAGGTGCGCAAGTGTGCAAGCTGTGCGCG 1513
Db 4181 ACTGGGCGACGCGATCTTACCCCGCATCAAGGTGCGCAAGTGTGCAAGCTGTGCGCG 4240
Qy 1514 GCGCCAAAGGCGCTGACCGACATCGTGCCCTGACCGAGAGGCGCTGAGAGCTGAGCGG 1573
Db 4241 GCGCCAAAGGCGCTGACCGACATCGTGCCCTGACCGAGAGGCGCTGAGAGCTGAGCGG 4300
Qy 1574 AGAACCAGAGATCTTGCGCGAGCGCGTGCACAGCGGTGTACTACGACCCCGACAAGACC 1633
Db 4301 AGAACCAGAGATCTTGCGCGAGCGCGCTGTCAGCGGTGTACTACGACCCCGACAAGACC 4360
Qy 1634 TGTGTGCGGAGATCCAGAAAGAGGCGCACAGCAAGTGAACCTTACGAGATCTTACGAGAGC 1693
Db 4361 TGTGTGCGGAGATCCAGAAAGAGGCGCACAGCAAGTGAACCTTACGAGATCTTACGAGAGC 4420
Qy 1694 CCTTCAAGACCTTGAAAGACCGGCAAGTACCGCAATCGCACCGGCGCCACCAACGACG 1753
Db 4421 CCTTCAAGACCTTGAAAGACCGGCAAGTACCGCAATCGCACCGGCGCCACCAACGACG 4480
Qy 1754 TGAAGCACTGACCGAGGCGGTGACAGAAAGATCGCATGTGAGAGCATCTGTGATCTTGGGCA 1813
Db 4481 TGAAGCACTGACCGAGGCGGTGACAGAAAGATCGCATGTGAGAGCATCTGTGATCTTGGGCA 4540
Qy 1814 AGACCCCAAGTTCCGCTGCTCCATCCAGAGAGAGACCTTGGAGACCTTGTGTGACCGACT 1873
Db 4541 AGACCCCAAGTTCCGCTGCTCCATCCAGAGAGAGACCTTGGAGACCTTGTGTGACCGACT 4600
Qy 1874 ACTGGCAGGCGCACTGAGATCCCGAGTGGAGTTGTGAACACCCCCCTGTGTGAAGC 1933
Db 4601 ACTGGCAGGCGCACTGAGATCCCGAGTGGAGTTGTGAACACCCCCCTGTGTGAAGC 4660
Qy 1934 TGTGTATCAAGCTGAGAGAGAGCCCATCATCTGCGCGCGAGACCTTCTTACGTGACGCGG 1993
Db 4661 TGTGTATCAAGCTGAGAGAGAGCCCATCATCTGCGCGCGAGACCTTCTTACGTGACGCGG 4720
Qy 1994 CCGCCACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGAGA 2053
Db 4721 CCGCCACCGCGAGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGCGCGCGAGA 4780
Qy 2054 AGATGTGTAAGCTTGACCGAGACCAACCAACCAAGAGAGTGTGACCGCATCTCCAGCTGG 2113
Db 4781 AGATGTGTAAGCTTGACCGAGACCAACCAACCAAGAGAGTGTGACCGCATCTCCAGCTGG 4840
Qy 2114 CCTTCAGAGACAGCGCGAGAGGTGAACATCTGTACCGACAGGCAAGTACGCGCTGGGCA 2173
Db 4841 CCTTCAGAGACAGCGCGAGAGGTGAACATCTGTACCGACAGGCAAGTACGCGCTGGGCA 4900
Qy 2174 TCATCAGGCGCGACCGCGACCAAGAGCGAGAGCTGTGAAACGAGTCAATCAGACAGC 2233
Db 4901 TCATCAGGCGCGACCGCGACCAAGAGCGAGAGCTGTGAAACGAGTCAATCAGACAGC 4960
Qy 2234 TGATCAAGAGAGAGAGGTGTACTCTGAGTGGTGCCCGCCCAAGAGGCGATCGCGCGCA 2293
Db 4961 TGATCAAGAGAGAGAGGTGTACTCTGAGTGGTGCCCGCCCAAGAGGCGATCGCGCGCA 5020
Qy 2294 ACAGACAGATCGACAGCTGTGAGAGAGGCGATCGCAAGGTGTCTTCTCGACGCGCA 2353
Db 5021 ACAGACAGATCGACAGCTGTGAGAGAGGCGATCGCAAGGTGTCTTCTCGACGCGCA 5080
Qy 2354 TCGATGCGCGCATCTGTATCTTACGAGTACATGAGACGACTGTACGTGTGCGACGCGCGC 2413
```

```
Db 5081 TCGATGCGCGCATCTGTATCTTACGAGTACATGAGACGACTGTAGTGGCAGCGCGCGC 5140
Qy 2414 CTAGATCGATTAAAGCTTCCCGGGCTTACACCGGT 2451
Db 5141 CTAGATCGATTAAAGCTTCCCGGGCTTACACCGGT 5178

RESULT 12
ACA03547
ID ACA03547 standard; DNA; 2457 BP.
XX
AC ACA03547;
XX
DT 22-MAY-2003 (first entry)
XX
DE Synthetic DNA encoding immunogenic HIV peptide #30.
XX
KW Immunogenic HIV polypeptide; human immunodeficiency virus; HIV; vaccine;
KW gene therapy; packaging cell line; humoral immune response;
KW cellular immune response; gene delivery vector; DNA immunisation; de.
XX
OS Synthetic.
XX
PN WO200304657-A1.
XX
PD 16-JAN-2003.
XX
PF 05-JUL-2002; 2002MO-US021421.
XX
PR 05-JUL-2001; 2001US-0303192P.
PR 31-AUG-2001; 2001US-0316860P.
PR 16-JAN-2002; 2002US-0349728P.
PR 16-JAN-2002; 2002US-0349793P.
PR 16-JAN-2002; 2002US-0349871P.
XX
PA (CHIR ) CHIRON CORP.
XX
PI Zur Megede J, Barnett SW, Lian Y;
XX
DR WPI, 2003-221602/21.
XX
PT New synthetic polynucleotides encoding antigenic HIV type B and/or type C
PT polypeptides, useful as immunogenic compositions or vaccines for
PT generating humoral or cellular immune responses against HIV in a subject,
PT especially humans.
XX
PS Example 1; Fig 35; 262pp; English.
XX
CC The invention describes a synthetic polynucleotide encoding 2 or more
CC immunogenic HIV polypeptides, where at least 2 of the polypeptides are
CC derived from different HIV subtypes. The polynucleotide is useful for
CC immunisation, generation of packaging cell lines, or production of HIV
CC polypeptides. The polynucleotide and its encoded proteins are useful as
CC immunogenic compositions or vaccines for generating humoral or cellular
CC immune responses against HIV in a subject, or for inducing neutralising
CC antibodies against HIV. The gene delivery vector comprising the
CC polynucleotide is also useful for DNA immunisation of, or for generating
CC an immune response (e.g. a humoral or cellular immune response) in, a
CC subject such as a mammal, particularly a human. This sequence encodes a
CC human immunodeficiency virus immunogenic peptide
XX
SQ Sequence 2457 BP; 564 A; 835 C; 758 G; 300 T; 0 U; 0 Other;

Query Match 98.8%; Score 2428.6; DB 8; Length 2457;
Beat Local Similarity 99.6%; Pred. No. 3.6e-292;
Matches 2447; Conservative 0; Mismatches 4; Indels 6; Gaps 1;

Qy 1 GTCCAGCGCACATGAGCGCGAGGCGATGAGCGACGAGCGCAATCTCTGATGCG 60
Db 1 GTCCAGCGCACATGAGCGCGAGGCGATGAGCGACGAGCGCAATCTCTGATGCG 60
Qy 61 CGACGCACTTCAAGGCGCGCAAGGCGATCAATGAGCTTCACTGCGGCAAGAGGCG 120
```

Dh 61 GCGAGCACTTCAAGGGCCCCCAAGGCGATCATCAAGTGTTCACACTGGCGCAAGAGGGC 120
Qy 121 CACATGCGCCGCACTGCGCGCGCCCCCGCAAGAGAGGCTGCTGCAAGTGGCGCAAGAG 180
Db 121 CACATGCGCCGCACTGCGCGCGCCCCCGCAAGAGAGGCTGCTGCAAGTGGCGCAAGAG 180
Qy 181 GGCACACAGATGAAGGACTGCAACGAGGCGCAAGGCGCACTTCTTCCGCGAGGACCTGGCC 240
Db 181 GGCACACAGATGAAGGACTGCAACGAGGCGCAAGGCGCACTTCTTCCGCGAGGACCTGGCC 240
Qy 241 TTCCCGCAAGGCAAGGCGCGCGAGTTCGCCAGAGCAGAACCGCGCCCAAGCCCCACC 300
Db 241 TTCCCGCAAGGCAAGGCGCGCGAGTTCGCCAGAGCAGAACCGCGCCCAAGCCCCACC 300
Qy 301 AGCCGCGAGCTGCAAGTGGCGCGCGCAACACCCCGCAGAGGCGCGCGCGCGCGCGCAG 360
Db 301 AGCCGCGAGCTGCAAGTGGCGCGCGCAACACCCCGCAGAGGCGCGCGCGCGCGCGCAG 360
Qy 361 GGCACCCCTGAACCTTCCCGCAAGTCAACCCCTGTGGCAGGCGCGCGTGTGACATCAAGGTG 420
Db 361 GGCACCCCTGAACCTTCCCGCAAGTCAACCCCTGTGGCAGGCGCGCGTGTGACATCAAGGTG 420
Qy 421 GCGCGGCAAGTCAAGAGGCGCGCTGTGACACCGGCGCGCAAGCAGACCGTGTGAGAGAG 480
Db 421 GCGCGGCAAGTCAAGAGGCGCGCTGTGACACCGGCGCGCAAGCAGACCGTGTGTGAGAGAG 480
Qy 481 ATGAGCGTGGCGCGCAAGTGAAGGCCCAAGATGATCGGCGGCACTGGCGGCTTTCATCAAG 540
Db 481 ATGAGCGTGGCGCGCAAGTGAAGGCCCAAGATGATCGGCGGCACTGGCGGCTTTCATCAAG 540
Qy 541 GTGGCGCAATTAAGACCAAGATCTGTATCGAGATCTGGGCGCAAGAGGCGCTCCGACCGTG 600
Db 541 GTGGCGCAATTAAGACCAAGATCTGTATCGAGATCTGGGCGCAAGAGGCGCTCCGACCGTG 600
Qy 601 CTGATGCGGCGCGCAACCCCGTGAACATCATCGGCGCGCAAGCTGTGACCGAGCTGGCTGC 660
Db 601 CTGATGCGGCGCGCAACCCCGTGAACATCATCGGCGCGCAAGCTGTGACCGAGCTGGCTGC 660
Qy 661 ACCCTGAACCTTCCCGCAAGTCAAGGCCCATCGAGACCGTGGCGCGTGAAGCTGAAGGCCGCGATG 720
Db 661 ACCCTGAACCTTCCCGCAAGTCAAGGCCCATCGAGACCGTGGCGCGTGAAGCTGAAGGCCGCGATG 720
Qy 721 GACGCGCGCAAGTGAAGAGTGGCGCGCTGACCGAGAGAGAAATCAAGGCCCTGACCGCC 780
Db 721 GACGCGCGCAAGTGAAGAGTGGCGCGCTGACCGAGAGAGAAATCAAGGCCCTGACCGCC 780
Qy 781 ATCTGGAAGAGATGAAGAGAGGCGCAAGTCAACCAAGTGGCGCGCGCGAGACCCCTTAC 840
Db 781 ATCTGGAAGAGATGAAGAGAGGCGCAAGTCAACCAAGTGGCGCGCGCGAGACCCCTTAC 840
Qy 841 AACACCCCGGTGTTCCGCAATCAAGAGAGAGACAGCACCAAGTGGCGCGCAAGCTGTGAGAC 900
Db 841 AACACCCCGGTGTTCCGCAATCAAGAGAGAGACAGCACCAAGTGGCGCGCAAGCTGTGAGAC 900
Qy 901 TTCCGCGAGCTGAACAGGCGCACCCGAGACTTCTGGAGGTGACGTGGCGCATCCCCCAG 960
Db 901 TTCCGCGAGCTGAACAGGCGCACCCGAGACTTCTGGAGGTGACGTGGCGCATCCCCCAG 960
Qy 961 CCGCGCGGCTGAAGAGAGAGAGAGGCGTGAACCGTGTGAGAGTGGCGCGAGCGCTTACTTC 1020
Db 961 CCGCGCGGCTGAAGAGAGAGAGAGGCGTGAACCGTGTGAGAGTGGCGCGAGCGCGCTTACTTC 1020
Qy 1021 AGCGTGGCGCTGACAGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGACATCAAC 1080
Db 1021 AGCGTGGCGCTGACAGAGACTTCCGCAAGTACACCGCGCTTCAACATCCCGACATCAAC 1080
Qy 1081 AAGAGAGACCCCGCGCATCCGCTACAGTACAGTGTGCGCCAGGCGCTGAGAGGCGCAGC 1140
Db 1081 AAGAGAGACCCCGCGCATCCGCTACAGTACAGTGTGCGCCAGGCGCTGAGAGGCGCAGC 1140
Qy 1141 CCGAGCATCTTCCAGAGAGCATGACCAAGATCTTGAAGCTTCCGCGCGCGCAACCCC 1200
Db 1141 CCGAGCATCTTCCAGAGAGCATGACCAAGATCTTGAAGCTTCCGCGCGCGCAACCCC 1200

Qy 1201 GAGATGTGATCTACAGGCGCGCGCGCTGTACGTTGGGCAAGCAGTGGAGATGGGCGACGAC 1260
Db 1201 GAGATGTGATCTACAGGCGCGCGCGCTGTACGTTGGGCAAGCAGTGGAGATGGGCGACGAC 1260
Qy 1261 CGCGCAAGATGAAGAGCTGGCGCAAGCAGTGTGCGTGGGCTTCAACACCCCGCAG 1320
Db 1261 CGCGCAAGATGAAGAGCTGGCGCAAGCAGTGTGCGTGGGCTTCAACACCCCGCAG 1320
Qy 1321 AAGAGCAGCAGAGAGAGCGCGCTTCTGCGCAT-----CGAGTGCACCCCGACAG 1374
Db 1321 AAGAGCAGCAGAGAGAGCGCGCTTCTGCGCATGGGCTTACGAGCTGCACCCCGACAG 1380
Qy 1375 TGGACCGTGGCGCGCATGAGTGGCGCGAGAGAGAGAGTGGACCGTGAACGACATCCAG 1434
Db 1381 TGGACCGTGGCGCGCATGAGTGGCGCGAGAGAGAGAGTGGACCGTGAACGACATCCAG 1440
Qy 1435 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGTCTACCCCGCATCAAGGTGGCGCAG 1494
Db 1441 AAGCTGTGGGCAAGCTGAACTGGGCGAGCGAGTCTACCCCGCATCAAGGTGGCGCAG 1500
Qy 1495 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCTGACCGAGAG 1554
Db 1501 CTGTGCAAGCTGTGCGCGCGCGCAAGGCGCTGACCGACATCGTGGCGCTGACCGAGAG 1560
Qy 1555 GCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTCCGCGAGCGCGTGCACGGCGCTTAC 1614
Db 1561 GCGGAGCTGGAGCTGGCGAGAACCGCGAGATCTTCCGCGAGCGCGTGCACGGCGCTTAC 1620
Qy 1615 TACGACCCCGAGCAAGACCTGTGGCGGAGTCCAGAACAGAGGCGCAAGCATGGTGAAC 1674
Db 1621 TACGACCCCGAGCAAGACCTGTGGCGGAGTCCAGAACAGAGGCGCAAGCATGGTGAAC 1680
Qy 1675 TACGAGTCTTACAGAGGCGCTTCAAGACCTGAAGACCGGCAAGTACGCAAGTGGCC 1734
Db 1681 TACGAGTCTTACAGAGGCGCTTCAAGACCTGAAGACCGGCAAGTACGCAAGTGGCC 1740
Qy 1735 ACCGCGCAACCAAGAGCTGAAGAGAGTGGCGCGAGCGGTGCAGAAATTCGCGATGAGAG 1794
Db 1741 ACCGCGCAACCAAGAGAGTGAAGAGAGTGGCGCGAGCGGTGCAGAAATTCGCGATGAGAG 1800
Qy 1795 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCGCGCTGCGCATCAAGAGAGACTGCG 1854
Db 1801 AGCATGTGATCTGGGCGCAAGACCCCGCAAGTTCGCGCTGCGCATCAAGAGAGACTGCG 1860
Qy 1855 GAGACCTGTGAGCGACTTACTGAGAGCGCACTGTGATCCCGAGTGGAGTTCGTGAAC 1914
Db 1861 GAGACCTGTGAGCGACTTACTGAGAGCGCACTGTGATCCCGAGTGGAGTTCGTGAAC 1920
Qy 1915 ACCCGCGCGCTGTGAAGCTGTGTATCAAGCTGAGAGAGAGGCCATCATCGCGCGCAG 1974
Db 1921 ACCCGCGCGCTGTGAAGCTGTGTATCAAGCTGAGAGAGAGGCCATCATCGCGCGCAG 1980
Qy 1975 ACCCTTCAAGTGAAGCGCGCGCGCAACCGCGAGACCAAGATGGGCAAGGCGCGCTACGTT 2034
Db 1981 ACCCTTCAAGTGAAGCGCGCGCGCAACCGCGAGACCAAGATGGGCAAGGCGCGCTACGTT 2040
Qy 2035 ACCGACCGGCGCGCGCAAGAGTGTGAGCTGACCGGACCAACCAAGAGAGAGCCGAG 2094
Db 2041 ACCGACCGGCGCGCGCAAGAGTGTGAGCTGACCGGACCAACCAAGAGAGAGCCGAG 2100
Qy 2095 CTGAGGCGCATCAAGTGGCGCTGACAGACCGCGAGCGAGGTGAACATCGTGAACGAC 2154
Db 2101 CTGAGGCGCATCAAGTGGCGCTGACAGACCGCGAGCGAGGTGAACATCGTGAACGAC 2160
Qy 2155 AGCGAGTACCGCTGGGCAATCAATCAAGCGCGCAAGCGCGCAAGAGCGAGCGAGCTGGT 2214
Db 2161 AGCGAGTACCGCTGGGCAATCAATCAAGCGCGCAAGCGCGCAAGAGCGAGCGAGCTGGT 2220
Qy 2215 AACGAGTATGAGAGAGCTGATCAAGAGAGAGAGTGTACTGTAGCTGGTGGCGCC 2274
Db 2221 AACGAGTATGAGAGAGCTGATCAAGAGAGAGAGTGTACTGTAGCTGGTGGCGCC 2280

QY 1141 CCCAGCATCTTCCAGAGCAGCATGAACCAAGATCTGAGACCCCTTCCGGCCCGCAACCCC 1200
DB 1141 CCCAGCATCTTCCAGAGCAGCATGAACCAAGATCTGAGACCCCTTCCGGCCCGCAACCCC 1200
QY 1201 GAGATCGTGAATCTACCAAGGCCCCCTGTACTGTGGCAGAGCACTTGAAGATCGGGCAGCAGC 1260
DB 1201 GAGATCGTGAATCTACCAAGGCCCCCTGTACTGTGGCAGAGCACTTGAAGATCGGGCAGCAGC 1260
QY 1261 CGCGCCAGAGATGAGAGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAGC 1320
DB 1261 CGCGCCAGAGATGAGAGAGCTGCGCAAGCACTGTGCGCTGGGGCTTCAACACCCCGCAGC 1320
QY 1321 AAGAGCAGCAGAAAGAGCCCCCTTCCCTGCCCAT-----GAGGTGAGACCCCGCAGCAG 1374
DB 1321 AAGAGCAGCAGAAAGAGCCCCCTTCCCTGCCCAT-----GAGGTGAGAGCTGACAGCTGACCCCAAG 1380
QY 1375 TGGACCGTGCAGACCCCATCGAGCTGCCCCGAGAGAGAGCTGGACCGTGAACGATCCAG 1434
DB 1381 TGGACCGTGCAGACCCCATCGAGCTGCCCCGAGAGAGAGCTGGACCGTGAACGATCCAG 1440
QY 1435 AAGCTGTGGGCAAGCTGAATCTGAGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCAG 1494
DB 1441 AAGCTGTGTGGGCAAGCTGAATCTGAGGCGCAGCCAGATCTACCCCGCATCAAGGTGCGCAG 1500
QY 1495 CTGTGCAAGCTGTGCGCGCGCCCGCAGAGCCCTGACCCGATATGTGTGCTTGAACCGAGAG 1554
DB 1501 CTGTGCAAGCTGTGCGCGCGCCCGCAGAGCCCTGACCCGATATGTGTGCTTGAACCGAGAG 1560
QY 1555 GCCGAGCTGAGACTGCGCCGAGAACCGGAGATCTGTGCGGAGCCCGTGCACCGCGCTGAC 1614
DB 1561 GCCGAGCTGAGACTGCGCCGAGAACCGGAGATCTGTGCGGAGCCCGTGCACCGCGCTGAC 1620
QY 1615 TACGACCCCGCAGAGAGCTGTGTGCGCAGATCTCAGAACAGGCGCCAGCAGCATGTGAGAC 1674
DB 1621 TACGACCCCGCAGAGAGCTGTGTGCGCAGATCTCAGAACAGGCGCCAGCAGCATGTGAGAC 1680
QY 1675 TACGAGATCTACCAAGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCCCAAGATGCC 1734
DB 1681 TACGAGATCTACCAAGAGCCCTTCAAGAACTGGAAGACCGGCAAGTACGCCCAAGATGCC 1740
QY 1735 ACCGCGCCAGCAGCAAGAGCTGTGAGAGAGCTGACCGAGCCGTGAGAGAAATGCCCATGTGAG 1794
DB 1741 ACCGCGCCAGCAGCAAGAGCTGTGAGAGAGCTGACCGAGCCGTGAGAGAAATGCCCATGTGAG 1800
QY 1795 AAGCATGTGATCTGAGGCAAGACCCCAAGTTCCGCTCCCATCCAGAACAGAGACCTG 1854
DB 1801 AAGCATGTGATCTGAGGCAAGACCCCAAGTTCCGCTCCCATCCAGAACAGAGACCTG 1860
QY 1855 GAGACTGTGTGAGCCGACTACTGTGAGAGCCACTGTGATCCCGAGTGGAGGTTCTGTAAC 1914
DB 1861 GAGACTGTGTGAGCCGACTACTGTGAGAGCCACTGTGATCCCGAGTGGAGGTTCTGTAAC 1920
QY 1915 ACCCCCCCTGTGTGAGAGCTGTGTGACTACAGCTGGAAGAAAGAGCCCATCATTCGGCGCGCAG 1974
DB 1921 ACCCCCCCTGTGTGAGAGCTGTGTGACTACAGCTGGAAGAAAGAGCCCATCATTCGGCGCGCAG 1980
QY 1975 ACCTTTACGTGAGACCGCGCCCGCAGACCGGAGAACCAAGATCGGCAAGGCGCGCTACGT 2034
DB 1981 ACCTTTACGTGAGACCGCGCGCGCAGACCGGAGAACCAAGATCGGCAAGGCGCGCTACGT 2040
QY 2035 ACCGACCGGCGCGCGGCAAGAGATGTGAGCTGAGCCGAGACCAACCAAGCAAGAGACCGAG 2094
DB 2041 ACCGACCGGCGCGCGGCAAGAGATGTGAGCTGAGCCGAGACCAACCAAGCAAGAGACCGAG 2100
QY 2095 CTGCAAGGCAATCCAGCTGCGCCCTGCGAGCAGGCGGAGGAGGTGAACATCTGTAACCGAC 2154
DB 2101 CTGCAAGGCAATCCAGCTGCGCCCTGCGAGAGACGCGGAGGAGGTGAACATCTGTAACCGAC 2160
QY 2155 AGCCAGTAGCGCCTGTGGCATCATTCAGGCGCCAGCCGAGCAAGAGCAGAGAGCTGTG 2214
DB 2161 AGCCAGTAGCGCCTGTGGCATCATTCAGGCGCCAGCCGAGCAAGAGCAGAGAGCTGTG 2220
QY 2215 AACCAAGATCATGAGAGCTGATCAAGAAAGAGAGAGGTGTAAGCTGTGAGCTGGGTGCCGCC 2274

DB 2221 AACCAAGATCATGAGAGCTGATCAAGAAAGAGAGGTGTAAGCTGTGAGCTGGGTGCCGCC 2280
QY 2275 CACAAAGGCGCATCGCGGCAAGCAGAGATCGAACAGCTGTGAGCAAGGCGATCCGCAAG 2334
DB 2281 CACAAAGGCGCATCGCGGCGCAAGCAGAGATCGAACAGCTGTGAGCAAGGCGATCCGCAAG 2340
QY 2335 GTGCTGTCTTCTGAGACGGCATGATGAGCGCGCATCTGATCTACCAATACATGAGACGACTG 2394
DB 2341 GTGCTGTCTTCTGAGACGGCATGATGAGCGCGCATCTGATCTACCAATACATGAGACGACTG 2400
QY 2395 TACGTGGCAGCGCGCGCCCTAGAGTCGATTAAGCTTCCCGGCGCTAGCACCGGT 2451
DB 2401 TACGTGGCAGCGCGCGCCCTAGAGTCGATTAAGCTTCCCGGCGCTAGCACCGGT 2457
RESULT 14
ABL39959
ID ABL39959 standard; DNA; 2469 BP.
XX
AC ABL39959;
XX
DT 15-MAY-2002 (first entry)
XX
DE Synthetic construct PR975(+) SEQ ID NO:30.
XX
KW Human immunodeficiency virus type C; antigenic HIV type C protein;
KW immunogenic; immunisation; gag; pol; vif; vpr; tat; rev; vpu; env; nef;
KW immunostimulant; gene therapy; gene; ds.
XX
OS Human immunodeficiency virus; type C.
XX
PN WO200204493-A2.
XX
PD 17-JAN-2002.
XX
PF 05-JUL-2001; 2001WO-US021241.
XX
PR 05-JUL-2000; 2000US-00610313.
XX
PA (CHIR) CHIRON CORP.
PA (UYST-) UNIV STELLENBOSCH.
PI Zur Megede J, Barnett SM, Engelbrecht S, Van Rensburg EJ;
PI MPI; 2002-154920/20.
DR
XX
PT New polynucleotides encoding antigenic HIV Type C polypeptides, useful in
PT applications including DNA immunization or generation of packaging cell
PT lines, particularly in gene therapy.
XX
PS Claim 1; Fig 8; 233p; English.
XX
CC The present invention describes expression cassettes comprising a
CC polynucleotide sequence encoding a polypeptide comprising immunogenic HIV
CC type C polypeptides. The expression cassettes comprise any of the HIV
CC type C sequences encoding Gag, Pol, Vif, Vpr, Tat, Rev, Vpu, Env or Nef
CC (I). (i) have immunostimulant activity and can be used in gene therapy.
CC The HIV type C polynucleotides are useful in applications including DNA
CC immunisation, generation of packaging cell lines, and production of HIV
CC type C proteins. The polynucleotides are particularly useful in gene
CC therapy and DNA immunisation applications. ABL39942 to ABL40054 and
CC ABB06204 to ABB06215 represent sequences used in the exemplification of
CC the present invention
XX
SQ Sequence 2469 BP; 571 A; 833 C; 761 G; 304 T; 0 U; 0 Other;
Query Match 98.3%; Score 2415.4; DB 6; Length 2469;
Best Local Similarity 99.3%; Pred. No. 1.5e-290;
Matches 2451; Conservative 0; Mismatches 6; Indels 12; Gaps 2;
QY 1 GTGACGCGCAGCATGGCGGCGCATGAGCCAGGCGCACAGCGGCCCAATCTGTGAGCAG 60

Db 1 GTGACGACCATGAGCGAGCCATGAGCCAGGCGCAACATCTGTATGAG 60
Qy 61 CGGAGCACTTCAAGAGGCGCCCAAGGCGATCATAGTCTTCAACTGAGCGCAAGAGGCG 120
Db 61 CGGAGCACTTCAAGAGGCGCCCAAGGCGATCATAGTCTTCAACTGAGCGCAAGAGGCG 120
Qy 121 CACATGCGCGGCACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGCAGAGAG 180
Db 121 CACATGCGCGGCACTGCGCGCGCGCGCGCAAGAGGCGTGTGAAAGTGCAGAGAG 180
Qy 181 GGCACACAGATGAAAGATGACACCGAGCGCCAGGCGCACTTCTTCGCGAGGACCTGCGC 240
Db 181 GGCACACAGATGAAAGATGACACCGAGCGCCAGGCGCACTTCTTCGCGAGGACCTGCGC 240
Qy 241 TTCCCGGAGGCGCAAGGCGCGGAGTTCCCAAGGAGCAAGACCGGCGCAACAGCGCCAC 300
Db 241 TTCCCGGAGGCGCAAGGCGCGGAGTTCCCAAGGAGCAAGACCGGCGCAACAGCGCCAC 300
Qy 301 AGCCGCGAGCTGCAAGTGCAGGCGCAACCCCGCGAGAGCGCGCGCGCGAGCGCGAG 360
Db 301 AGCCGCGAGCTGCAAGTGCAGGCGCAACCCCGCGAGAGCGCGCGCGCGAGCGCGAG 360
Qy 361 GGCACCTGAACTTCCCGCAGATCACTCTGTGAGCGCGCGCGCTGTGTGAGCATCAAGGTG 420
Db 361 GGCACCTGAACTTCCCGCAGATCACTCTGTGAGCGCGCGCGCTGTGTGAGCATCAAGGTG 420
Qy 421 GCGCGCGAGATCAAGAGGCGCGTGTGCAACCGGCGCGAGACACCGTGTGTGAGAGAG 480
Db 421 GCGCGCGAGATCAAGAGGCGCGTGTGCAACCGGCGCGAGACACCGTGTGTGAGAGAG 480
Qy 481 ATGAGCCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540
Db 481 ATGAGCCTGCGCGGCAAGTGAAGCCCAAGATGATCGCGCGCATCGCGCGCTTCAATCAAG 540
Qy 541 GTGCGCGAGTACAGACCAAGTCTGTATCGAGATCTGCGCGAGAGAGGCGCATCGCGCAC 600
Db 541 GTGCGCGAGTACAGACCAAGTCTGTATCGAGATCTGCGCGAGAGAGGCGCATCGCGCAC 600
Qy 601 CTGATCGCGCGCGCGCGCGTGAACATCATCGCGCGCGCAACATCGCGCGCGCGCGCG 660
Db 601 CTGATCGCGCGCGCGCGCGCGTGAACATCATCGCGCGCGCGCAACATCGCGCGCGCGCG 660
Qy 661 ACCCTGAACCTTCCCGCATCAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAGCG 720
Db 661 ACCCTGAACCTTCCCGCATCAGCGCGCGCATCGAGACCGTGCCTGTGAAGCTGAGCG 720
Qy 721 GACGCGCGCGGAGTGAAGTGTGCGCGCTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
Db 721 GACGCGCGCGGAGTGAAGTGTGCGCGCTGACCGAGAGAGAGATCAAGGCGCTGACCGCG 780
Qy 781 ATCTGCGAGGAGATGAGAGAGGAGGAGATCAACAGATGCGCGCGCGGAGACCGCTAC 840
Db 781 ATCTGCGAGGAGATGAGAGAGGAGGAGATCAACAGATGCGCGCGCGGAGACCGCTAC 840
Qy 841 AACACCGCGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGCAGCGCATGTGTGAGC 900
Db 841 AACACCGCGTGTGCGCATCAAGAGAGAGAGACAGACCAAGTGCAGCGCATGTGTGAGC 900
Qy 901 TTTCGCGAGCTGAACAGCGCACCCAGAGCTTCTGTGAGGTGCACTGTGAGCATCCCGAC 960
Db 901 TTTCGCGAGCTGAACAGCGCACCCAGAGCTTCTGTGAGGTGCACTGTGAGCATCCCGAC 960
Qy 961 CCGCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGTGAGAGCGCGCTACTTC 1020
Db 961 CCGCGCGCGCTGAAGAGAGAGAGCGTGAACCGTGTGAGAGTGTGAGAGCGCGCTACTTC 1020
Qy 1021 AGCGTGCCTGTGAAGAGAGCTTCTGCAAGTACACCGCTTCAACCATCCCGACGATCAAC 1080
Db 1021 AGCGTGCCTGTGAAGAGAGCTTCTGCAAGTACACCGCTTCAACCATCCCGACGATCAAC 1080
Qy 1081 AACGAGACCCCGCGCATCGCTTACAGATCAACGTGTGCGCGCGAGGCTGAGAGGCGAGC 1140
Db 1081 AACGAGACCCCGCGCATCGCTTACAGATCAACGTGTGCGCGCGAGGCTGAGAGGCGAGC 1140

Db 1081 AACGAGACCCCGCGCATCGCTTACAGATCAACGTGTGCGCGCGAGGCTGAGAGGCGAGC 1140
Qy 1141 CCGAGCATCTTCAAGAGGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGACCGC 1200
Db 1141 CCGAGCATCTTCAAGAGGAGATGACCAAGATCTGTGAGCGCTTCCGCGCGCGACCGC 1200
Qy 1201 GAGATCGTATCTACCA-----GCGCGCGTGTACGTGAGCGAGCATGTGAGATCGAGC 1254
Db 1201 GAGATCGTATCTACCAAGATCATGAGAGACCTGTACGTGTGAGCGAGCATGTGAGATCGAGC 1260
Qy 1255 CAGCACCGCGCGCAAGATGAGAGAGTGTGAGAGACCTGTGCGTGTGAGCGCTTACACAC 1314
Db 1255 CAGCACCGCGCGCAAGATGAGAGAGTGTGAGAGACCTGTGCGTGTGAGCGCTTACACAC 1320
Qy 1315 CCGGACAAAGAGACACAGAGAGAGCGCGCTTCTGTGCGCAT-----GAGCTGACCGC 1368
Db 1315 CCGGACAAAGAGACACAGAGAGAGCGCGCTTCTGTGCGCAT-----GAGCTGACCGC 1380
Qy 1369 GACAAAGTGAACCTGTGACAGCGCGCATGAGCTGCGCGAGAGAGAGAGAGCTGAGACGAGC 1428
Db 1369 GACAAAGTGAACCTGTGACAGCGCGCATGAGCTGCGCGAGAGAGAGAGAGCTGAGACGAGC 1440
Qy 1429 ATTCAGAGAGCTGTGTGAGAGAGTGAACCTGTGAGAGCTGAGAGCTGAGAGCTGAGAG 1488
Db 1441 ATTCAGAGAGCTGTGTGAGAGAGTGAACCTGTGAGAGCTGAGAGCTGAGAGCTGAGAG 1500
Qy 1489 GCGCAGCTGTGACAGCTGTGTGCGCGCGCGCAAGGCGCGTGAACGACATGCGCGCGCTGAC 1548
Db 1501 GCGCAGCTGTGACAGCTGTGTGCGCGCGCGCAAGGCGCGTGAACGACATGCGCGCGCTGAC 1560
Qy 1549 GAGAGGCGGAGCTGTGAGAGCTGTGCGCGAGAGACCGGAGATCTGTGCGAGAGCGCGTGCAGCG 1608
Db 1561 GAGAGGCGGAGCTGTGAGAGCTGTGCGCGAGAGACCGGAGATCTGTGCGAGAGCGCGTGCAGCG 1620
Qy 1609 GTGTACTGACACCCGAGAGAGAGCTGTGTGCGCGAGATCTGAGAGAGAGAGAGAGAGAG 1668
Db 1621 GTGTACTGACACCCGAGAGAGAGCTGTGTGCGCGAGATCTGAGAGAGAGAGAGAGAGAG 1680
Qy 1669 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGACCTGAGAGACCGGCGAGATGAGCGCAAG 1728
Db 1681 TGAGACTTACAGATCTTACAGAGAGCGCTTCAAGAGACCTGAGAGAGCGGCGAGATGAGCGCAAG 1740
Qy 1729 ATGCGCACCGGCGCACACCAAGAGAGAGAGCTGAGCGAGCGCGTGCAGAGATGCGC 1788
Db 1741 ATGCGCACCGGCGCACACCAAGAGAGAGAGCTGAGCGAGCGCGTGCAGAGATGCGC 1800
Qy 1789 ATGAGAGAGATGTGTGATCTGTGCGCGCAAGACCGCGCAAGTTCCGCTGCGCATCAAGAGAG 1848
Db 1801 ATGAGAGAGATGTGTGATCTGTGCGCGCAAGACCGCGCAAGTTCCGCTGCGCATCAAGAGAG 1860
Qy 1849 ACTGAGAGAGCTGTGTGAGACGAGCTTACAGAGAGCGCTGAGATCCCGGAGTGGAGTTTC 1908
Db 1861 ACTGAGAGAGCTGTGTGAGACGAGCTTACAGAGAGCGCTGAGATCCCGGAGTGGAGTTTC 1920
Qy 1909 GTGAAACACCGCGCGCTGTGTGAGAGCTGTGTGATCAAGCTGAGAGAGAGAGCGCATCTGCGC 1968
Db 1921 GTGAAACACCGCGCGCTGTGTGAGAGCTGTGTGATCAAGCTGAGAGAGAGAGCGCATCTGCGC 1980
Qy 1969 GCGGAGACCTTCTTACGTGTGAGAGCGCGCGCGCAACCGCGAGACCAAGATGCGAGAGCGCGC 2028
Db 1981 GCGGAGACCTTCTTACGTGTGAGAGCGCGCGCGCGCAACCGCGAGACCAAGATGCGAGAGCGCGC 2040
Qy 2029 TACGTGACCGAGCGCGCGCGCGCGAGAGATCGTGAAGCTGAGACCGAGACCAACCAAGAG 2088
Db 2041 TACGTGACCGAGCGCGCGCGCGCGAGAGATCGTGAAGCTGAGACCGAGACCAACCAAGAGAG 2100
Qy 2089 ACCGAGCTGAGAGCGCATCAAGCTGCGCGCTGAGAGAGAGAGAGAGAGAGTGAACATCTGTG 2148
Db 2101 ACCGAGCTGAGAGCGCATCAAGCTGCGCGCTGAGAGAGAGAGAGAGAGAGTGAACATCTGTG 2160
Qy 2149 ACCGAGCGCGAGTACCGCTGTGAGCATCATCGAGCGCGCGCGAGACCAAGAGAGAGAGAG 2208
Db 2161 ACCGAGCGCGAGTACCGCTGTGAGCATCATCGAGCGCGCGCGAGACCAAGAGAGAGAGAGAG 2220

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 30, 2005, 07:51:41 ; Search time 8676.48 Seconds
(without alignments)
13249.138 Million cell updates/sec

Title: US-09-610-313B-32

Sequence score: 2457
1 gtcgacgcacacatgcccga.....gggctgacacggcgaatc 2457

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 2339354128 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_esc1:*
2: gb_esc2:*
3: gb_esc3:*
4: gb_esc4:*
5: gb_esc5:*
6: gb_esc6:*
7: gb_esc7:*
8: gb_esc8:*
9: gb_esc9:*
10: gb_esc10:*
11: gb_esc11:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	136.4	5.6	330	10	CL293849
2	100	4.1	2886	10	CL967755 OaIFCC015
3	90.2	3.7	1781	8	CKX09623 RRCm0260
4	90.2	3.7	2598	4	AY103647 Zea mayb
5	88.8	3.6	3069	10	CL973991 OaIFCC025
6	86.6	3.5	1398	10	CL961989 OaIFCC006
7	86	3.5	743	10	CZ247380 AIAA-aaf3
8	85.6	3.5	951	3	BM321451 rockefell
9	84.2	3.4	1941	10	CL971508 OaIFCC021
10	80.8	3.3	1060	10	CW922203 EDGAR29TR
11	80.6	3.3	869	7	CK159167 FAS04056
12	80.2	3.3	892	10	CZ216254 AIAA-aaf2
13	80.2	3.3	1132	3	BM320864 rockefell
14	79.6	3.2	1165	3	BM320900 OaIFCC049
15	79.6	3.2	1680	10	CL982270 OaIFCC049
16	79.6	3.2	867	3	BM321430 rockefell
17	78.4	3.2	1725	10	CL978463 OaIFCC031
18	78.4	3.2	2031	10	CL974989 OaIFCC042
19	78.2	3.2	2697	10	CL952258 OaIFCC000
20	78.2	3.2	2853	10	CL974397 OaIFCC025
21	78	3.2	1485	10	CL970981 OaIFCC020
22	77.4	3.2	1509	10	CL959255 OaIFCC002

23	77.4	3.2	2559	10	CL982027 OaIFCC046
24	77	3.1	11691	10	CL962901 OaIFCC008
25	76.8	3.1	2682	10	CL968033 OaIFCC017
26	76.6	3.1	1550	3	BM321022 rockefell
27	76.2	3.1	1401	10	CL962721 OaIFCC038
28	76.2	3.1	3249	10	CL945510 OaIFCC004
29	75	3.1	545	3	BM724851 103107580
30	75	3.1	862	3	BM321023 rockefell
31	74.4	3.0	914	9	BZ568300 pac82-164
32	74.2	3.0	2313	10	CL982362 OaIFCC047
33	74	3.0	1290	10	CL972679 OaIFCC023
34	74	3.0	1386	11	DQ045165 Homo sapi
35	74	3.0	2072	4	CR603312 full-1eng
36	73.8	3.0	2433	10	AY401196 Homo sapi
37	73.2	3.0	853	3	BM321393 rockefell
38	72.8	3.0	566	3	BM587428 170006873
39	72.6	3.0	718	4	CNS06HPN Tetradon
40	72.6	3.0	788	8	CB643171 OSJNB03L
41	72.4	2.9	753	9	CC675888 OGMCO51TH
42	72.4	2.9	2523	10	CL974879 OaIFCC026
43	72.2	2.9	602	7	CV057146 BNE124e3
44	72.2	2.9	640	2	BE601575 HVSMB009
45	72.2	2.9	688	6	CB648640 OSJNB12C

ALIGNMENTS

RESULT 1
LOCUS CL293849/330 bp DNA linear GSS 12-FEB-2004
DEFINITION 0280349-08A1-C03 UniformMu MutRIL Library Zea mays genomic clone
ACCESSION CL293849
VERSION CL293849.1 GI:42541978
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 330)
Lathaw,S., Tan,B.-C., Settles,A.M. and McCarty,D.R.
Sequence tagged transposon insertions from the UniformMu maize population
Unpublished (2003)
JOURNAL Contact: Donald R. McCarty
Plant Molecular and Cellular Biology Program
University of Florida
PO 110690 Gainesville, FL 32611-0690, USA
Tel: 352-392-1928 x322
Email: drmcufl.edu
Sequence flanking probable Mu insertion site in UniformMu line:
0280349-08, Primer set: A
Class: transposon insertion site.
FEATURES
source
1..330
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="W22 (ACR, bz1-m)"
/cultivar="UniformMu"
/db_xref="taxon:4577"
/clone="0280349-08A1-C03"
/note="Vector: TOPO-PCR4; DNA flanking Mu transposon insertions in Mu inactive lines were extracted from the UniformMu maize population by the thermo asymmetric interlaced PCR (TAIL) protocol using primers specific for the Mu terminal inverted repeat and a set of 16 arbitrary primers. Amplicons were size enriched using Sephadex 400 spin columns and cloned into the TOPO PCR4 vector."

ORIGIN

Query Match	Similarity	5.6%	Score 136.4	DB 10	Length 330
Best Local	Similarity 66.4%		Pred. NC.3.1e-14		
Matches 211	Conservative 0		Mismatches 106		Indels 1
			Gaps 1		
QY	421	GGCGGCCGATCAAGAGGAGCCCTGTCTGACACCGCGCGCCGACGACACCGTGTGAGANG	480		
Db	329	GGGGGGCAGCTGGAAGAGCTCTATTAGTATACAGGACAGATGATATACGATTATTAAGAA	270		
QY	481	ATGAGCCTTCGCCCCGAGAGTGGAAAGCCCAAGATGATCGCGCGCATCGCGGCTTTCATCAAG	540		
Db	269	ATGACTTTTGAACAGGAAGATGGAAACCAAAAAGATAGGGGGAATTGGAGTTTATCAA	210		
QY	541	GTGCGCCAGTACGACACAGATCCTGATTCGAGATCTCGCGCAGAGAGGCCATCGGCACCGTG	600		
Db	209	GTTAAGACAGTATGATCAAGTACCGTACCGATAGAAATCTGTGGCATTAAGCTATAGGTACGTTA	150		
QY	601	CTGATTCGCGCCCAACCCCGGTGAATCATTCGCGCCGCAACATGCTGACCCAGCTCGGCTGC	660		
Db	149	TTAGTAGAGACCTACACCTGTCAACTAATTTGGAAGAAATCTGTTGACTCGAGTTGG-TCG	91		
QY	661	ACCCTGAACCTTCCCATCATGACCCCATCGAACCGTGGCCCGTGAACCTGAAGCCCGGCATG	720		
Db	90	ACCTTAATTTTCCCATTAGTCTATTTGAATCTGTACAGTAAATTTAAAGCCAGGAATG	31		
QY	721	GACGGCCCCAAGGTGAAG	738		
Db	30	GATGGCCCAAAAGTAAAG	13		

RESULT 2	LOCUS	DEFINITION
CL967755	2886 bp	DNA
CL967755	2886 bp	linear
CL967755	2886 bp	GSS
CL967755	2886 bp	21-SEP-2004
CL967755	2886 bp	Oryza sativa Express Library Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.

VERSION	KEYWORDS	SOURCE	ORGANISM
CL967755.1	GI:52390149	GSS.	
			<i>Oryza sativa</i> (indica cultivar-group)
			<i>Oryza sativa</i> (indica cultivar-group)
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

AUTHORS	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis
JOURNAL	Unpublished (2004)
COMMENT	Contact: Chen Chang

Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel.: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

```

SOURCE
1. 2886
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="caxon:39946"
/clone_lib="Oryza sativa Express library"
/note="Oryza sativa exon tripped genomic sequences"

```

[illegible]

Db	219	GGTATATGAGAGCGCGCGCATATATGCGCATATGCTCTCGCAACGAGCGGCGGACAGCGCGC	278
Oy	249	GGGCGAAGCGCCGCGAGTTTCCCACGCGAGCAAGACCGCGCAACAGCCCTCACACGCGCGCA	308
Db	279	CGACTGCGAGGACTTTCGTTCGCGCATCTGTCCTCTCTGTCATCAATCTCACACATCAAGTTT	338
Oy	309	GCTGCAAGTGGCGC--GGCGCAACCCCGGAGCGAAGCGCGGCGCGAGCGCGAC	365
Db	339	CATGAGAGAAACATATCCGGAAACGCGCGCGCGGCGCTCATATGCGCGGCTTCGCGCCCA	398
Oy	366	CTTGAATTTCCCGCATATCACTCTGTGGACGCGCCCTGTGTAGCATCAAGTGGCGG	425
Db	399	GACCAAGTGTCTAGGGACCGGAAATGGCAGAGACAGACGCGGTCATCTCGTTCCGG	458
Oy	426	CCAGATCAAGAGGCGCTGCTGACACCGGCGCGCATGACACGCTGCTGAGAGAGATGAG	485
Db	459	CGACATCATGACATCAAGCTCGCGCACATCATCCCGCGACGCGCGGCTGCTCGAGGG	518
Oy	486	CTTGGCGCGGCAATGGAAGCCCAAGATGATCGGCGGATCGGCGGCTTCATCAAGTGGCG	545
Db	519	CGACCTCGCTCAAGTTCGACCAAGCGCGCTTCAAGCGAGTTCGATGCGCTTCACAAAGCA	578
Oy	546	CCAATGACACAGATCTCTGATTCGATCTCGCGCAAGAGGCTCATCGGCAACGCTGAT	605
Db	579	CGCGGGGACGGGCGCTTCTTCGGGCTCACCGTCAACAGGCGAATTCGAGCGCGTGT	638
Oy	606	CGGCGCCACCCCGGTGAACATCATCGGCGGACATGCTGACCCGCTGAGCTGACCTT	665
Db	639	CATGCGCACCGGCGTGGCACACTTCTTCGCGCAAGCGGCGCACCTGTGTGACAGCACAA	698
Oy	666	GAACTTCCCATGAGCCCATCGAGACCGTGGCCGTTGAGAGCTGAAGCCCGGCAATGACGG	725
Db	699	CAACATGGCGACTTTCAGGCTCGTGTCTCAGGCGATCGGCACTTTCGATCATCTTCAT	758
Oy	726	CCCCAAGGTGAAGCAGTGGCGCCCTCGACCGAGAGAGATCAAGGCGCTTGAACCGCATCTG	785
Db	759	CGGCGTGGGATATATTCAGAGATCATGTTATGTATCCGATTCAGACACGCGCGTACCG	818
Oy	786	CGAGAGATGAGAGAGGCGCAAGATGACCAAGATTTGGCGCGGAGAACCCCTTACAAAC	845
Db	819	CGAGGGATCGACAACTGCTCGTCTGCTCATCGGCGGATCCCACTGCGCATGCGCAC	878
Oy	846	CCCCGTTCGCGCATCAAGAAAGGACAGACCAATGGCGCAAGTGGATGACTTCCG	905
Db	879	CTGCTGTGCTGACCATGCGCATCGGCTTCCACCGGCTGTTCGAGAGAGGCGCATAC	938
Oy	906	CGAGCTGAACAAGGCGACCCAGGACTTCTGGAGGTTCAGCTGGGCAATCCCGCACCCCGC	965
Db	939	CAACGCGATGACCGCGCATCGAGAGATGGCGCGCATGACGTGTCTGACGCGACAAAGAC	998
Oy	966	CGGCGTTAAGAAAGAAAGAGCGGTGACCTGTGTGGAAGTGGGCGAGCGCTTACGCGT	1025
Db	999	CGGAGCGCTGACGCTCAACAAGCTCACCGTTCGACAAACCTGTGATCGAGGTGTTCGGGCG	1058
Oy	1026	GCCCCCTGAGAGAGACTTTCGCGACGTACCTGCTTACCATCCCGCATCAACAAAGCA	1085
Db	1059	GGGCGTGGACAGGACTCGGTGCTCTGTACGCGCGAGGCGGTCCGCGTCAAGAAACA	1118
Oy	1086	GACCCCGGACCTCGGCTACCAAGTACCAAGTACGATCCCGCAAGGCGAGGCGCCAG	1145
Db	1119	GGAGCGCATTTGACATGATCTGTGGGCACTGTTCGCGACCCCAAGAGGCGCGCGCG	1178
Oy	1146	CATCTTCAGAGCAGATGACCAAGATCTTGAAGCCCTTTCGCGCGCGCAACCCCGAGAT	1205
Db	1179	CATCAAGAGGTTCACCTTCTC-----CGTTCAACCCGTCGAGAAAGCGAC	1226
Oy	1206	CGTATCTTACAGGCGCCCGCTGTATCGTGGGCAAGGACCTTGAGATTCGCGCAGACCGCGC	1265
Db	1227	GGCCATCACCTTACATGACGCGCAATGGCGAATGCGACAGATTCGACAAAGGCGCGCGGA	1286
Oy	1266	CAAGATCGAGGCGTGGCGCAAGCAGCTGTGCGCTTGGGGCTTTCACCAACCCCGCAAGAA	1325
Db	1287	GCAATCATCAAGCTGTGTGCAAGATGAGACCAAGAGCGCGCAAGAAAGAGGTGACACGCTGAT	1346

Qy	1326	GCACCAAGAGAGGCCCCCTTCCTCGGCCCATGAGGTGCACCCCGACAAGTGAACGTGCA	1385
Db	1347	CGACCACTGATCCCGAACCGCGGCTTCCTCGGTGCGTGTGTTCAACGAAGTGTCCGGA	1406
Qy	1386	GCCCATCGAGCTGCGCCGAGAAGAGAGCTGAGCCGTGACACGATCCAGAGCTGTGTGG	1445
Db	1407	GAAAGACGAAGAGAGCGAGGGCGAGGCCATGGCAGTTGCTCGGCTCTCTGCGGCTGTTGGA	1466
Qy	1446	CAAGCTGAATCTGGGCGACCGACAGATCTTACCCCGGCATCAAGGTGCGCAGCTGTGCAGCT	1505
Db	1467	CCCCCGAGGCAACGACGACGCGGAGAACATCCGCGCGCTGACACTCGCGTGAAGT	1526
Qy	1506	GCTCGCGGCGCGCCCAAGCGCTTGAACCGACATGTGTGCTCCCTTGACCGAGAGGCGGAGCTGGA	1565
Db	1527	GAAATGATCTACCGCGCGACCAAGCTTCGCATCG-----GGAAGGAGACGCGGCGCGGCT	1580
Qy	1566	GCTGGCGAGAACCGCGAGATCTGTGCGGACCGGTGACGCGCGGTGTACTACACCCAG	1625
Db	1581	CGGATGGGACCAACATGTAACCGGTGACGACGCTGTGGCGACCAAGACAGAGAGAT	1640
Qy	1626	CAAGGACCTGTGTGGCGGATCTCAAGACAGGCGCACGACGATGGAACCTTACAGATTA	1685
Db	1641	GAGCGGCTTCCCATCTCAACGAGCTGATGGAAGAGCGCACGGCTTCGCGCGCGTGTCC	1700
Qy	1686	CCAGAGGCCCTTCAAGAACTTGAAAGACCGGCAAGTACGCAAGATGCGACCGGCCACAC	1745
Db	1701	GGAGCACAAGTACGAGATCTGTAGCGCTTCAAGACCGAAGACCATCTGCGGCTATAC	1760
Qy	1746	CAAGCAGCTGAAGCAGCTGACCGAGGCGGTGCAAAAGTCCGCATGAGAGAGATCTGTAT	1805
Db	1761	CGGCGACGCGCTGAAAGACGACGCGCGCGCTCTGAAGAGCGCAATCGGCATCTGCGTGA	1820
Qy	1806	CTGGGCGAAGACCCCC	1821
Db	1821	CGACGCAACGACCGC	1836

RESULT	3
CX099623	
LOCUS	1781 bp mRNA linear EST_03-JUN-2005
DEFINITION	RcOM260 A normalized whole-life-cycle cDNA library of rice Oryza sativa (Indica cultivar-group) cDNA clone E1057014, THREEB09, E1057004, E1062D05, B1027I14, E1070G16, E108 5', mRNA sequence.
ACCESSION	CX099623
VERSION	CX099623.1 GI:66912775
KEYWORDS	EST.
SOURCE	Oryza sativa (indica cultivar-group)
ORGANISM	Oryza sativa (indica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrenfeldiae; Oryzaceae; Oryzae. 1 (bases 1 to 1781) Zhang,J., Feng,Q., Jin,C., Qiu,D., Zhang,L., Xie,K., Yuan,D., Han,B., Zhang,Q. and Wang,S. Features of the expressed sequences revealed by a large-scale analysis of ESTs from a normalized cDNA library of the elite indica rice cultivar Minghui 63 Plant J. 42 (5), 772-780 (2005) 15918889
TITLE	Contact: Wang S
JOURNAL	National Key Laboratory of Crop Genetic Improvement
PUBMED	Hauzhong Agricultural University
COMMENT	Wuhan 430070, China Tel.: 86-27-87282044 Fax: 86-27-87287092 Email: shiyingwang@hotmail.com Seq primer: T7.

```

FEATURES
source      location/Qualifiers
1. 1781
   /organism="Oryza sativa (indica cultivar-group) "
   /mol_type="mRNA"
   /strain="indica"

```

```

/cultivar="Minghui_63"
/db_xref="taxon:39946"
/clone="E1057014, T3REB09, E1057004, E1062D05, B1027114,
E1070616, E108"
/tissue_type="whole plant"
/dev_stage="whole-life-cycle"
/lab_host="E. coli DH10B"
/clone_1fb="A normalized whole-life-cycle cDNA library of
rice"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
library is constructed based on the strategy of saturation
hybridization with genomic DNA using rice cultivar Minghui
63. This library consists of cDNA from 15 directionally
cloned cDNA libraries constructed with different tissues
from 9 developmental stages."

```

Query Match	3.74;	Score 90.2;	DB 8;	Length 1781;
Best Local Similarity	43.84;	Pred. No. 6	5e-06;	
Matches 447;	Conservative 0;	Mismatches 568;	Indels 6;	Gaps 1;

Query Match	3.7%	Score 90.2	DB 8	Length 1781
Beet Local Similarity	43.8%	Pred. No. 6.5e-06		
Matches 447	Conservative	0	Mismatches 568	Indels 6
			Gaps	1
QY	231	GAACTTGACCTTCCCCCAGGGCAAGGCGCGGAGTTCCCGACGACGAGAACCGCGCCAA	290	
DB	194	GAGACGGGTGCTGACACCGGTGCTCGCGCCAGAACCCCGAGCAAGGATGGCTTGGAGAC	253	
QY	291	CAGCCCCACAGCGCGGAGCTGCAAGTGGCGGGGACAAACCCCGCAGCGAGGCGCGCGC	350	
DB	254	GTGCAACCAAGACCAACATGTGTATGTGTGTTCGGGAGATCAACCAAGGCAACCTGTGA	313	
QY	351	CGAGCGCAGGGGACCCCTGAACCTTCCCCCGAATCAACCTGTGGGAGAGGCCCTCGTAG	410	
DB	314	CTTACGAGAAAGATCTGTGGCGACACTTCCGCGGATGGCTTCTGTTCGAGACGTCGG	373	
QY	411	CATCAAGATGGCGCGGCAGATCAAGAGAGCGCTTGCTGACACCGCGCGACGACCGT	470	
DB	374	CCTGAGAGCGGACCGCGTGAAGGTGTGTCTCAATCATGAGACGACGATCGCCGACATCGC	433	
QY	471	GCTGAGAGAGATGAGCTTGCCTGGCGCAAGTGAAGCCCAAGTATGTGGCGGCATCGCGG	530	
DB	434	GCAAGGGGGTGCACGGCCACTTTCACCAAGCGCCCGAAGAGATTCGGCGCGCGGACGAGG	493	
QY	531	CTTCATCAAGGTGGCGCAGTACGACGACGAGATCTGATGAGATCTTGGCGCAAGAGGCAT	590	
DB	494	CCAATATTTGGGTCAACGCAACCGACGAGAACCCCGAGCTGATGGCCCTCAGCCACTGTCT	553	
QY	591	CGGACCGGTGTGATCGGCGCCACCCCGGTGAACATATCGGCGGCAACATGCTGACCCA	650	
DB	554	CGCACCAAGTTCGGCGCGCGCGCTCACCGAGAGTTCGGCAAGAACGGGCACTTGGCGGTGCT	613	
QY	651	GCTGGGGTGCACCTTGAACTTCCCATCAAGCCCATGGAAGACCGTGGCCCGTGAAGCTGAA	710	
DB	614	CAGGCGCGGACGGCAAGAACCCAGGTACCGTGAAGTCTTCAACGACGCGCGCGCATGTGT	673	
QY	711	GCCTGGGATGAGCGGCGCCCAAGGTGAGAGAGTGGCGCCCTGACCGAGAGAAATCAAGGC	770	
DB	674	CCCGGTCCGGTCCACACACCGTCTCATCTTCAACCGACGAGAGAGAACCGTCAACCAACGA	733	
QY	771	CCTGACCGCATCTTGGAGAGATGAGAGAGAGAGGCAATCAACCAAGATCGCGCCCGCA	830	
DB	734	CGAATGCGCGCGC-----GACCTCAAGAGAGCAGTCAATCAAGCGCGGTCAATCCCGACAA	787	
QY	831	GAAACCCCTCAACAACCCCGGTGTGGCATCAAGAAAGGACAGCAACCAAGTGGCGGCA	890	
DB	788	GTACTCTGACAGAAAGACATCTTCAACTCAACCCCTCGGCGCGCTTCTGTCAATCGGCGG	847	
QY	891	GCTGTGTGACTTTCGCGAGCTGAAACAAGCGCACCAAGACTTCTGGAGGTGACGCTGGG	950	
DB	848	GCCCCAGCGACGCGCGGCTTCAACCGCGCGAAGTCAATCATGACACTTACGCGCGATG	907	
QY	951	CATCCCCCAACCCCGCGGCTTGAAGAAAGAAAGAAAGAGCTGACGCTGTGAGCTGTGGGGA	1010	
DB	908	GAGGCGGACGCGCGCGCGCTTCTTCGCGCAAGAACCCGACCAAGGTGACCGGAGCGG	967	

Db	Accession	Version	KeyWords	Source	Organism	Reference Authors	Journal Title	Journal Pubmed	Journal Reference Authors	Journal Title	Journal Comment	Features Source
Qy	1011	CGCCTACTTCAAGCGTGGCCCTTGAGACGAGGATCTTCGGAAGTACACACCGCTTACACATGCC										
Db	968	CGCCTTCAATCCGACAGGACAGGCGCGCCAAAGACATGTCGCCACGAGCGCTCGCCGCGCTG										
Qy	1071	CAGCATCAACACGACGAGACCCCGGCGATCCGCTACCAAGTACACAGTGCCTGCCCCAGGACTG										
Db	1028	CATGTGTGAGAGTGTGTATCGCATTCGCGATTCGCGGCTCCCGAGCGCGTCTTCGGTGTTCGCACTC										
Qy	1131	GAAAGGACAGCCCCAGCATCTTCCAGAGCAGCATGACCAAGATCCTGAGGCCCTTCCGCGC										
Db	1088	CTACGCGCACCGGCAAGATCCCCGACAGAGAGATCCTCAAGATCGTCAAGAGAACTTCGCA										
Qy	1191	CCGCAACCCCGAGATCGTATCTACGAGCGCCCTGTATAGTGGGACGCGACCTGAGAT										
Db	1148	TTTCAGGCGCCGGAGTATGATACCATCACTGACCTCAAGAGGCGGCAACCGGTTTCAT										
Qy	1251	C 1251										
Db	1208	C 1208										
RESULT 4	AY103647											
LOCUS	AY103647	2598 bp	mRNA	linear	HTC 18-FB8-2005							
DEFINITION	Zea mays PC0142084 mRNA sequence.											
ACCESSION	AY103647											
VERSION	AY103647.1	GI:21206725										
KEYWORDS	HTC.											
SOURCE	Zea mays											
ORGANISM	Zea mays											
REFERENCE	Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta, Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae, PACCAD clade, Panicoidae, Andropogoneae, Zea.											
AUTHORS	1 (bases 1 to 2598) Gadiner, J., Schroeder, S., Polacco, M. L., Sanchez-Villeda, H., Fang, Z., Morgante, M., Landave, T., Fongler, K., Usache, F., Hanfey, M., Tingey, S., Chou, H., Wing, R., Soderlund, C. and Coe, E. H. Jr.											
TITLE	Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization											
JOURNAL	Plant Physiol. 134 (4), 1317-1326 (2004)											
PUBMED	15020742											
REFERENCE	2 (bases 1 to 2598) Hainey, C. F., Dolan, M., Miao, G. H., Vogel, J. M., Whitesalt, M. S., Arthur, L. W., Hanfey, M., Morgante, M. and Tingey, S. V.											
AUTHORS	Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes											
TITLE	Unpublished (2002)											
JOURNAL	3 (bases 1 to 2598)											
REFERENCE	Coe, E. H.											
AUTHORS	Direct Submission											
TITLE	Submitted (25-Apr-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA											
JOURNAL	Missouri, Columbia, MO 65211, USA											
COMMENT	If you are interested in getting corresponding physical clones, these are publicly available from ZmDB and may be found by BLAST searching at MSU, maizegap.org; ZmDB, www.zmdb.iastate.edu; TIGR, www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the maize cDNA sequences is either Virginia Walbot, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZmDB: www.zmdb.iastate.edu.											
FEATURES	Location/Qualifiers											
SOURCE	1..2598											
	/organism="Zea mays"											
	/mol_type="mRNA"											
	/db_xref="MaizeGDB:638378"											
	/db_xref="taxon:4577"											
	/clone_lib="Maize Mapping Project/DuPont Consensus Library"											
	/note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed DuPont contigs; this resource was assembled by DuPont as part of a collaboration for the											

overgo addressing of BACs in conjunction with the Maize Mapping Project"

Query Match	Similarity	3.7#	Score 90.2	DB 4	Length 2599
Best Local	Similarity	42.2#	Pred. No. 6.7e-06		
Matches	814	Conservative	0	Mismatches 1096	Indels 15
				Gaps	5
QY	CCACCAGCCCGCAGACTGACAGGTGCGCGCGCAACCCCGCAGCGAGGCGCGCCGAGC	296			355
Db	CCACCACCTCCCACTCTGCGCACCGCGCGCCCAACCAACACACCGCAGCGCGACAA	51			110
QY	GCGAGGGACCCCTGAATTTCCCCAGATCACTCTGTGCAACGCCCTTGTGTGAGATCA	356			415
Db	TTGGCGGGGAAAGGCGCCATCTGTGAGAGCGACCCGCTGAATCTGGGCGCGCGCGCG	111			170
QY	AGGTGGCGCGCAATCAAGAGGCGCCTGTGACACCGCGCGCGACGACACCGTCTG	416			475
Db	AGCTGGCGGGAGCGCACTGTGACAGAGTGAAGCGCATGTGTGCCCGCAGCGCCG	171			230
QY	AGGAGATGAGCTGCGCCCGGCAATGTGAAGCCCAAGATGATCGCGCGCATCGCGCTTCA	476			535
Db	TGTCTCAAGATGAGAGGTTCACCTCTCGCGTGTGCGCAGGTGCGCGCGCTCCCTCCGCA	231			290
QY	TCAAAGTTCGCGCAATGACACCAATTCCTGATCTGAGATCTGTGGCGCAAGAGGCGCATCGGCA	536			595
Db	AGGACGGCTCGCGCGTCCCGCTGAGCTCGACGAGAGAGGCGCGCCCGCTCAAGGCGCA	291			350
QY	CCGTGCTGATCGCGCCCAACCCCGGTGACATCATCGCGCGCAATGCTGACCCGACTG	596			655
Db	GCAACGATGATCTTGACATGCAATGCGCCACCGCGGGGACATCTACGGCGTCAACACG	351			410
QY	GCTGACCCCTGAATCTTCCCATCAAGCCCATTCAGACCGGTGACCTGTAAGTCAAGCCCG	656			715
Db	GCTTCGGGGGACCTTCCACCGCGCGCACCAAGAGCGGGCCGCGCTTCAGGTTCGAGTGC	411			470
QY	GCATGACCGGCCCCAAGGTGAAGAGTGAAGGCCCCCTGACCGAGAGAGAGATCAAGGCCCTGA	716			775
Db	TCAGGCACTCAAGCGCGGAATCTTTCGACACCGGCAAGGACAGGACACACGCTCGGTGG	471			530
QY	CCGCCATCTGCGAGAGATGAGAGAGGCGCAAGATCAACCAAGATCGGCCCGGAGAAC	776			835
Db	AGGTACCCCGCGCGCGCATGTCTGTGCGCATCAACCTCTCCACAGGCTTACCTCCGCA	531			590
QY	CCCTCAACACCCCGGTGTTTGGCCATCAAGAGAGAGAGACAGACCAAGTGGGCGAACGTGG	836			895
Db	TCGGCTTGAGATCTCGAGGCGCATCACAGAACTGCTCAACACCGGTGTACGCCCTTCCG	591			650
QY	TGGACTTCCGCGAGCTGAACAAGCGCACCCAGGACTTCTGGAGAGTGAAGCTGGGATCC	896			955
Db	TGCGGCTTCGGGGGACATCAACCGGTTCGGGACGCTGGTCCCGCTCTCTACATTCGCG	651			710
QY	CCCAACCCCGCGGCTTGAAG	956			1015
Db	GCTCATCATCGGCGCGGCCCAACGCGCAGGCGCTTCAACG---	711			767
QY	ACTTCAGGTCGCCCTGTGACGAGGACTTCGCAAGTACACCGCCTTCAACATCCACAGA	1016			1075
Db	ACGCGCGCGAGAGGCTTCAAGATCGCGGATCGAGGCGGCTTCTTCAAGCTCAACCCCA	768			827
QY	TCAAACAAGAGACCCCGGCGCATTCGCTACAGTACCAAGTGTGTCGCCCGAGGAGTGAAGG	1076			1135
Db	AGGAGGGCGTGCATGTGTCAACCGGCAAGTCCGTGGGCTCGCGCTCGCGGCAACCGTGA	828			887
QY	GCAAGCCCGAGATCTTTCAGAGCAGCATGACCAAGATCTTGAAGCCTTCCGCGCGCGCA	1136			1195
Db	TGTACAGACGCAAGTCTCTGCGCGTCTGTGTGAGGTCT---	888			944
QY	ACCCCGAGATGTATCTTACAGAGGCGCCCTGTACGTGTGGGAGAGGACTTGAAGATCGAGC	1196			1255
Db	AGGTATATAGAGGAGACGCCGAGTACAGGACACACTGACCCACAGCTTGAAGACACACC	945			1000
QY	AGCACCGGCGCAAGATCGAGAGGCTGTGCGAACACTGTGCTG---	1256			1312

Db	1005	CGGGGTCAATCGAGGCCGCGGCCCATCATGAGGACATCTTGGAATGGGAGCTCTTCAATGA	1064
QY	1313	CCCCGACAGAGAGCACAGAGAGAGGCCCTTCTCTGCCCATGAGCTGACCCGACAC	1372
Db	1065	AGCAGGGCCAGAAATTGAAACAGAGCTGAGCCCGCTGTGGAAGCCCAACAGGACAGGTATCG	1124
QY	1373	AGTGAACCGTGCACGCCCATGAGGTGCGCCGAGAGAGAGAGAGTGAACCGTGAACGACATCC	1432
Db	1125	CGCTCCGACGTGCGCCGACGATGAGGCTTGAGGCCCGCCAGATGAGGTATTCGCGCGCCACCA	1184
QY	1433	AGAACTGTGTGGGCAAGCTGAATCTGGGCGACCGACGATCTACCCCGGACATCAAGGTGCGCC	1492
Db	1185	AGTCCATTCGAGCGGAGGGTCAACTCCGTAACGA---CAACCCGGTATCGACGTTCACCC	1241
QY	1493	AGCTGTGGAAGTGTCTGTGCGCGCGCCCAAGGCCCTGACCGACATGTGTCCCTTGACCGAGG	1552
Db	1242	GGCGCAAGGCGCTGACGCGCGGCAACTTCCAGGGGCAACCCCATGCGGTGTCCATGACAC	1301
QY	1553	AGGCGGAGCTGGAAGCTGGCCGAGAACCGCGAGATCTTGCGCGAGCCCGGTGCACCGCGGTGT	1612
Db	1302	AAGCGCGGCTGGCATGCGCATGCGCAACATCGGCAAGCTCATATTTGCGGAGTTCTTCGAGCTCG	1361
QY	1613	ACTACGACCCCAAGAGAGACTGTGTGCGCGAGATTCAGAGACAGGGGCCACGACATGTGGA	1672
Db	1362	TCAACGAGTTTACAAACAACGCGGTCTCACTTCAACCTGGCCGCGACGCGCAACCCGACC	1421
QY	1673	CCTACCGAGATTCACAGGAGGCCCTTCAAGAACCTGAAGACCGGCAAGTACGCCCAAGATGC	1732
Db	1422	TGGACTACGGCTTTCAGAGGGCACCGAGATGCGCATGAGCTCTTACTGTCTCGAGCTCCAGT	1481
QY	1733	GCACCGGCCCAACCAACGACGTGAAGACGTGAACCGAGGCCGTGCAGAGATGCGCCATGG	1792
Db	1482	ACCTGGGCAACCCCATCAACAACGCGGACAGACGCGCGAGACACACAAACAGGACGTGA	1541
QY	1793	AGAGCAAT---CGTGAATCTGGGGCAAGACCCCGCAAGTTCGCGCTGCCATTCAGAAAGAGA	1849
Db	1542	ACTCCCTGAGGCTGTGCTCGGACAGAAAGACCGCGAGGCGATGACATCTCTGAAGCTCA	1601
QY	1850	CCTGGGAGACCTGTGTGGAACCGACTTACTGGCAGGCCACCTGTGATCCCGAGTGGGAGTTG	1909
Db	1602	TGTGTGCCACTTACATCGGTGGCGGTGTGCCAGGCGGTGTGAACCTTGCGCCACTCGAGGAGA	1661
QY	1910	TGAACACCCCCCTCTGTGTAAGCTGTGTACAGCTGTGAAGAGAGGCCCATCATCGGCG	1969
Db	1662	ACATTCMAAGCGTGTGTGAAGAACACCGTGAACCCAGGTGGCCAGAAAGGTGTGTAACATGA	1721
QY	1970	CCGAGACCTTTCACGTGAGCGGCGCGCGCAACCGGAGAACCAAGATCGGGCAAGGCCGCT	2029
Db	1722	ACCCCTCGGGGAGAGCTTCTCAAGCGCCGCTTCAAGGAGAGAGAGTATCATAGCGGCATCG	1781
QY	2030	ACGTGAACCGACCGGGGCGGCGAGAAAGATCTGAAGCTGAGCCGTAACCGAGACCAACAACGAAGA	2089
Db	1782	ACCAGGAGGCGGTGTTTCACTGAACGGGAGAGACGCGGACAGGCGGACGTGCGCGTGAATGC	1841
QY	2090	CCGAGCTGCACAGGCATTCAGACTGTGGCCCTGCAGAGACAGGGGACGAGGTGAACATCTGGA	2149
Db	1842	AGAACTCGCGCGCGTGTGTGGTGAACACAGCCCTCAGACAGGCGGACACCGGAGCGGGAGC	1901
QY	2150	CCGACAGCCAGTACGCGCTGTGGCATCATTCAGGCCCAAGCCGACCAAGACGAGAGCGACG	2209
Db	1902	CCTCGGTGTTCTCCAAAGTCAACAGATTCAGAGAGAGAGACTCGCGCGGTGTGCGCCAGG	1961
QY	2210	TGTTGAA 2216	
Db	1962	AGGTGA 1968	

RESULT 5	CL973991	LOCUS	DEFINITION
	CL973991	3069 bp	DNA
			linear
			GSS 21-SEP-2004
			OsIFCC025416 <i>Oryza sativa</i> Express Library <i>Oryza sativa</i> (indica cultivar-group) genomic, genomic survey sequence.

Accession	Version	Source	Keywords	Organism	Reference Authors	Title	Journal Comment	Features	Origin
CL973991.1	GI:52402507	GSS.	Oryza sativa (indica cultivar-group)	Oryza sativa (indica cultivar-group)	Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Zhao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis	Unpublished (2004)	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.	Location/Qualifiers 1..3069
Query Match	3.6%;	Score 88.8;	DB 10;	Length 3069;					
Best Local Similarity	41.5%;	Ident. No. 1.2e-05;							
Matches	652;	Conservative 0;	Mismatches 917;	Indels 3;	Gaps 1;				
Oy	37	ACGAGCCCAACATCTGATGACAGCGACGACAACTTCAAGGGCCCCAAGCGCATCATCAAG	96					/organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_id="Oryza sativa Express Library" /note="Oryza sativa exon trapped genomic sequences"	
Db	625	AGCGGCATCGGCAAGACGACGCTGGCAGCTACTCTTACCGAAGAACAGATCCGAG	684						
Oy	97	TGCTTCAACTGCGGCAAGAGAGGACACATGCGCCGCACTGCGCGCCCCCGCAAGAG	156						
Db	685	CATTGCGCTGCAAGCGGTGGGTTCAGTCCCGCAAGTTCGCGCTTCGCGCATCTCTC	744						
Oy	157	GGCTGCTGGAAGTGGCGGCAAGAGAGGCGCCAGATGAAAGATGCACTGCAACGACGCGCC	216						
Db	745	CAGCAGCTCATCGGCGCAGGCCAACGCTCTCGCGGACGACGAGCTTGGCGGAGGCG	803						
Oy	217	AATCTTCTCGGAGGACCTTGGCTTCCCGGAGGGCAAGGCCCCGAGTTCCCGAGCAG	276						
Db	804	--CGCCGACGACGACGTCGCGCGCGCGGGGTGGATGAGAGGCGAGAGGCTGTCAAGCG	861						
Oy	277	CAGAACCGCGCAACAGCCCCACCAAGCCCGGAGCTGACAGTGGCGGGGACCAATCCCGC	336						
Db	862	CGCTTCGCGCGCGGCTTCAGAGGGAAGAGTAACTCTGTGTCTGACGACGTCCGAGC	921						
Oy	337	AGCAGGCGCGCGCGGAGCGCCAGAGGCACTTCCCGGAGATCAACCTGTGGAG	396						
Db	922	AGGAGGAGTGGGCTTCTTCTCTGCGCGGCTTCCCGCGGCTTCGCGGAGCTGTGCTG	981						
Oy	397	CGCCCTCTGTGACATCAAGTGGGCGGCGACATCAAGAGGCGCTGTGAGACACCGGC	456						
Db	982	CTGCTCAAGACGACGATCAAGTCAAGTCAAGAGAGTCCGAGCAACCGTGGCGCGCGGCG	1041						
Oy	457	GCCTGACGACCGTGTGAGAGAGATGAGCTTGGCCCGCAAGTGGAGGCCCAATATATC	516						
Db	1042	GGCTGTGTCGACCAAGGAGCTCGGAAAGCTGTGGCGCGAGCAGCGCGCGCACTGTTC	1101						
Oy	517	GGCGGATCGGCGCTTATCAAGTGTGCGCAGTACGACGATCTTGATGATCTGC	576						
Db	1102	CGCGGAGGCTGTACGCGCACCGGAGCGCGGAGCGCGAAGAGTGAAGCAGCTCAAGGC	1161						

QY 577 GCGAAGAAAGCCATCGGCAACCGTCTGATTCGGCCCAACCCCGGTGAACATCATCGGCGC 636
 DB 1162 CTCGTGAGCTCATGACGAAGGGCTTCAACCTGCGCTCAACATGTCATGTCGCGGA 1221
 QY 637 AACATGTCAGCCAGCTGGCTGACCCCTGAACCTTCCCATGAGCCCATGAGACCGTG 696
 DB 1222 CTCCTGAGGTCCMAAAGAGATGATGAGGCGCTGATTCATAGCTTCATGATACG 1281
 QY 697 CCCGTGAAGCTGAAGCCCGGCAATGAGCGGCCCAAGGTGAAGCATGCGCCCTGACGAG 756
 DB 1282 CCAACCCCGGACATCGCCGGAAGACGCGCCGAGGCGAGCCCAAGAGACAGCTTC 1341
 QY 757 GAGAAATCAAGGCGCTGACCCGCTGTCGAGAGAGATGAGAGAGAGAGAGAGATCAAC 816
 DB 1342 GATGATAGAGAGAGCGCGCGGATCATGACGAGAGCAAGAGAGAGAGAGAGAGAG 1401
 QY 817 AAGATCGGCGCGGAGAACCCCTTCAACACCCCGTGTTCGTCATGAGAGAGAGACAC 876
 DB 1402 CAGATGAGGACATTCATGAGCAAGATTAATGACGATGTCATGAGACGACCTGCGCAG 1461
 QY 877 ACCAATGCGCGCAAGCTGGTGAACCTTCGCGAGCTGAACAAGCGACCCAGGACTTTCG 936
 DB 1462 CTGAAGCGGTGCTTCTTCTTCTTCTGCGCGGTTTCAACCGCGGACAGCCGATCTGCGCG 1521
 QY 937 GAGGTGACGCTGGGATCCCGCACCCCGCGCTGAGAGAGAGAGAGAGAGAGAGAG 996
 DB 1522 AAGCTGTCCGCTGTGGGTGCGCGAGGGTTTATGACGAGGAGAGAGAGAGAGAG 1581
 QY 997 CTGGAAGTGGGAGAGGCTTCTTCAAGCTGCTTGAAGAGAGAGAGAGAGAGAGAG 1056
 DB 1582 GAGGAGCAAGGAGAGAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1641
 QY 1057 GCTTCAACATCCCGAGATCAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1116
 DB 1642 GAG 1701
 QY 1117 CTGCGGAG 1176
 DB 1702 TTGCTGAG 1761
 QY 1177 GAGGCTTTCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1236
 DB 1762 CTCTCAAG 1821
 QY 1237 AGGCACTGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1296
 DB 1822 GTCATGCTGAG 1881
 QY 1297 CGCTGGGCTTCAACACCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1356
 DB 1882 GCGAG 1941
 QY 1357 GAGCTGACCCCGAG 1416
 DB 1942 CCCACTTCTCGTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2001
 QY 1417 ACCGGAAGAGATCAAG 1476
 DB 2002 CTGCGGACCGGAG 2061
 QY 1477 GGCATCAAG 1536
 DB 2062 GGCCTTGTGAG 2121
 QY 1537 GTGCGCTGAG 1596
 DB 2122 CAG 2181
 QY 1597 CCGGTGACGAG 1608
 DB 2182 CCTGCTCGC 2193

RESULT 6
 CL961989
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Oryza sativa (indica cultivar-group)
 Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatoidae; Oryzaceae; Oryza.
 1 (bases 1 to 1398)
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Xiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W., and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 JOURNAL
 COMMENT
 CONTACT: Chen Chen
 Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 Tel: 86-10-80481559
 Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.
 Location/Qualifiers
 1..1398
 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Expressed Library"
 /note="Oryza sativa exon trapped genomic sequences"
 ORIGIN
 Query Match 3.5%; Score 86.6; DB 10; Length 1398;
 Best Local Similarity 45.9%; Pred. No. 2.8e-05;
 Matches 508; Conservative 0; Mismatches 569; Indels 30; Gaps 5;
 QY 1 GTGACGCGCACATGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
 DB 166 GTGACGAG 225
 QY 61 GCGAGCACTTCAAG 120
 DB 226 GCGGCGGAG 282
 QY 121 CAGATGCGGCGGAG 180
 DB 283 CGATGATGCGGCGGAG 342
 QY 181 GAGCAACGATGAAG 240
 DB 343 GTCAAG 402
 QY 403 TTCCAGATCTCTCAACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 462
 DB 241 TTCCCGGAG 300
 QY 301 AGCGGAG 360
 DB 463 GCGGCGGAG 513
 QY 361 GGCACCTGGAATTTCCCGAGATCACTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 514 GCGCTGAG 570
 QY 421 GCGGCGGAG 480
 DB 571 ACACCAACCTGTGTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 630

QY 481 ATGAGCTGCTCCCGGCAATGAGAGCCCAAGATGATCGGGGATCGGGCTTCATCAAG 540
 DB 631 AGGATCTGCTCGCTGATCCCAATGCTCGGCTCTGGAGAGCCACCGGCGCATCGTCCG 690
 QY 541 GTGGCCCAATGAGAGCAAGATCTGATCGA---GATTCGGGCAAGAGGCAATCGGCAAC 597
 DB 691 GAGGTGCTGCTTACCGCGGTGGCAAGAGCTCAGAGATCGGGCCAAAGTTTCTCAACGCGAC 750
 QY 598 GTGCTGATCGGCCCCACCCCGTGAACATCATCGGCGGCAAGATGCTGAACCAAGCTGGG 657
 DB 751 GTGGGATTCGGTGGCTCTGCTTCCAGAAAGACATCTCACTGCTG----- 798
 QY 658 TGGACCTGAACTTCCCATCAAGCCCATCGAGACCTGTCCTGGTAAAGTGAAGCCCGAC 717
 DB 799 TACATCTGCGAGTGCATGAGCTCTCCCGAGGCGCACTACTGAGAAAGAGTTATCAAG 858
 QY 718 ATGAGCGGCCCCAAGGTGAAGAGAGTGGCCCTGACCGAGAGAAATCAAGGCGCTGACC 777
 DB 859 ATCAACGACTTACCAAGAGCGCGTTCGTGAACCGGCTGTGTCTTCATGTTCAACACCC 918
 QY 778 GCCATCTGCGAGAGATGAGAGAGGCAAGATCAACCAAGATCGGCCCCGAGAACCC 837
 DB 919 GTGGCGCGCAAGAGATGCGCGGTGCTGGCTTCGCTTCAAGAGAGACACCGGCGACAC 978
 QY 838 TACAAACACCCCGTGTGTTCCCATCAAGAGAGAGAGACACCAAGTGGCGCAAGCTGTG 897
 DB 979 AAGGAGACCGCCGCTCATCACTGCAAGGCGCTGATCGCGCAAGGCGCAAGGTTGAC 1038
 QY 898 GACTTCGCGAGGTGAACAGGCGACCTTCGTGGAGGTGAGCTGGGCAATCCCC 957
 DB 1039 ATCTACGATCCCCAGGTGACGAGAGACAGGTTCAGCGGACCTCGCCATGAGCAAGTTT 1098
 QY 958 CACCCCGCGGCTGAG 1017
 DB 1099 GATTTGGATTCACCGGCTCATCTCCAGCGAGTGAAGCCCAACGCGCATCAAGAGGTGTC 1158
 QY 1018 TTTCAGGTGCGGCTGAG 1077
 DB 1159 GTTCGCTGAG 1218
 QY 1078 AACACAG 1104
 DB 1219 TGGAG 1245

RESULT 7
 C2247380 743 bp DNA linear GSS 11-FEB-2005
 LOCUS DEFINITION A1AA-aat35a12.b1 Ancylostoma caninum whole genome shotgun library
 ACCESSION (A1AAGSS 001) Ancylostoma caninum genomic, genomic survey sequence.
 C2247380
 VERSION C2247380.1 GI:59632821
 KEYWORDS GSS
 SOURCE Ancylostoma caninum (dog hookworm)
 ORGANISM Ancylostoma caninum
 BUKARYOZA; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
 Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
 1 (bases 1 to 743)
 REFERENCE Miteva,M., McCarter,J.P., Pape,D., Rittler,B., Tsagaris,I.I., R.,
 AUTHORS Ronko,I., Martin,J., Wylie,T., Dante,M., Meyer,R., Messina,D.,
 Waterston,R.H., Clifton,S.W., and Wilson,R.
 TITLE Genome Survey sequences from the parasitic nematode Ancylostoma
 caninum
 JOURNAL Unpublished (2004)
 COMMENT Contact: Miteva M
 Washington University in St. Louis
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: nematode@watson.wustl.edu
 Genomic DNA provided by John Hawdon (mtnjmh@wumc.edu) DNA

Sequenced by Washington University Genome Sequencing Center
 Class: shotgun.
 FEATURES
 Location/Qualifiers
 source
 1..743
 /organism="Ancylostoma caninum"
 /mol_type="genomic DNA"
 /strain="Baltimore"
 /db_xref="taxon:29170"
 /dev_stage="adult"
 /lab_host="GSI0"
 /clone_lib="Ancylostoma caninum whole genome shotgun
 library (A1AAGSS 001)"
 /note="Vector: pOTW13; Site 1: BstXI; Site 2: BstXI;
 Ancylostoma caninum genomic DNA was randomly sheared,
 end-repaired and size fractionated to enrich for 2-4 kb
 fragments. Genomic DNA was provided by John Hawdon
 (mtnjmh@wumc.edu) at George Washington University.
 Sequencing by Washington University Genome Sequencing
 Center, St. Louis, MO."

ORIGIN
 Query Match 3.5%; Score 86; DB 10; Length 743;
 Best Local Similarity 45.1%; Pred. No. 3.5e-05;
 Matches 320; Conservative 0; Mismatches 390; Indels 0; Gaps 0;
 QY 777 CCGCATCTGCGAGAGATGAG 836
 DB 5 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 64
 QY 837 CTACAAACACCCCGTGTGTTCCCATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
 DB 65 CGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 124
 QY 897 GGAATTCGCGAGCTGAACAGCGGACCTTCGTGGAGGTGAGCTGGGCAATCCC 956
 DB 125 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 184
 QY 957 CACCCCGCGGCTGAG 1016
 DB 185 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 244
 QY 1017 CTTGAGCGGCGGCTGAG 1076
 DB 245 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 304
 QY 1077 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1136
 DB 305 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 364
 QY 1137 CAGCCCGAGATCTTTCAG 1196
 DB 365 CGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 424
 QY 1197 CCGCGAGATCGATCTACAG 1256
 DB 425 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 484
 QY 1257 GCAACCGCGCAAGATCGAG 1316
 DB 485 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 544
 QY 1317 CGACAAG 1376
 DB 545 CGACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 604
 QY 1377 GACCGTGAAG 1436
 DB 605 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 664
 QY 1437 GCTGTGGGCAAGCTGAAGTGGGCGAGAGATTAACCCGCGATCAAG 1486
 DB 665 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 714

RESULT	8
LOCUS	BM321451
DEFINITION	BM321451 951 bp mRNA linear EST 03-JAN-2002 rockefeller.0.1211 Masticigameoba balamuthi lambda ZAP II Library Masticigameoba balamuthi cDNA similar to adenovirahomocysteinease (EC 3.3.1.1), mRNA sequence.
ACCESSION	BM321451
VERSION	BM321451.1 GI:18055857
KEYWORDS	EST.
SOURCE	Masticigameoba balamuthi
ORGANISM	Masticigameoba balamuthi
REFERENCE	Eukaryota; Pelobiontida; Masticigameobidae; Masticigameoba. 1 (bases 1 to 951)
AUTHORS	Bajcskei, R., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Dunfee, L., Gaasterland, T., Lopez, P., Muller, M., and Philippe, H.
TITLE	The analysis of 100 genes supports the grouping of three highly divergent amebae: Dictyostelium, Entamoeba, and Masticigameoba
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
PUBMED	11830664
COMMENT	Contact: Muller Miklos Laboratory of Biochemical Parasitology The Rockefeller University 1230 York Avenue, New York, NY 10021, USA Email: mmuller@rockvax.rockefeller.edu Insert Length: 951 Std Error: 0.00 POLYA-No.
FEATURES	Location/Qualifiers
Source	1..951 /organism="Masticigameoba balamuthi" /mol_type="mRNA" /strain="ATCC 30984" /db_xref="taxon:108607" /clone_lib="Masticigameoba balamuthi lambda ZAP II library" /note="Syn: Phreatamoeba balamuthi"
ORIGIN	
Query Match	3.5% Score 85.6; DB 3; Length 951;
Best Local Similarity	46.0%; Pred. No. 4.2e-05;
Matches 323; Conservative	0; Mismatches 376; Indels 3; Gaps 1;
Dn	1722 CGCCAGATGCGCACCGGCCAACCAACGCTGAAGACGTGACCAGGCCTGTGCAGAA 1781
Dn	219 CGCCAGATGCGCTGTGCTGTGTCATATTTCTCGACGACGACACACGGCGCGCGC 278
Dn	1782 GATGCGCATGAGAGACATCTGTATCTTGGGGCAAAGCCCCCAAGTTCCGCTCCCATCA 1841
Dn	279 CATGCGCAGAGCGCGCGGTCTGTGCTTGCGCTGGAAGGCGCAAGAACTTCAGAGACTCG 338
Dn	1842 GAAGAGACCTGGAGAGACTGTGTGACCGAATCTACTGACGAGGCCACTGTGATCCCGAGTG 1901
Dn	339 GAGTGTGACCTGTGAAGGCCCTGTGCTTGTGGCCCCCTTACAGAGGCCCTCAGATCATGTGTGA 398
Dn	1902 GAGTGTGTGAAACACCCCCCTGTGTGAAGCTGTGTGTAACAAGCTGAGAAGAGAGCCCAT 1961
Dn	399 CGACGCGCGGTGACGCGCACTGTGATGATCCAAGAGGGGTTTGGGGCCGAGCAACCCCA 458
Dn	1962 CATGCGGCCCGAGACCTTCTATAGTGTGAAGGGCCCGCCCAACCGCGAGACCAAGATGTGGCAA 2022
Dn	459 GCTGTGTGAGAGACGACGAGAGGCTCTGAGAGAGGTGTGCTGTCAACAAGTGTCTTAAGCA 518
Dn	2022 GGCGCGGCTACGTGACCGAACCGGGGCGCGCAGAAAGATGTGNAAGCTGTGACCGAGACCAACCA 2083
Dn	519 GGTTCAGAGAGAGACGCCCGGTTTGTGGCACAAGATCTTCCCAGAGATTCGGGGTGTCHG 578
Dn	2082 CCAGAGAACCGAGCTGCAGGGCCATCAACTGTGGCCCTGTGACAGACAGCGGCAAGGAGTGA 2141
Dn	579 CGAGGAGACGACGACTGGCGTTATAGAGCTGTACCAAGCTGCACCGCGACGGCAAGCTGT 638
Dn	2142 CATGTTGACCGACACCAAGTAGTATCCCTTGGGACATCATCAAGGCTCCAGCCGACAGAGAGCA 2201
Dn	639 GTTCCCGGCGCTGCAAGCTCAAGAC--CTTWTGACCAAGAGCAAGTTGTGACAAATCTTA 695

Oy	2202	GAGCGAGCTGGTGAACCAAGATCATCTGAGCAAGCTATCAAGAAAGAAAGGTGTACTCTAG	2267
Db	696	CGGCTGCCTCCACATCTCGCTCATCTGACAGGCATCAAGCGAGCGACCAAGCTGAATGCTCGAGCG	755
Oy	2262	CTGGGTGTCGCCCCCAAGAGGCATCTGGGGGCAACGAGAGATTCGACCAAGTGGTGAAGCA	2322
Db	756	CAAGTGCCTCTCTCGCGGGCTACAGCGAGAGCTGGGCAAGAGGCTCGCGCGAGTGCCTCG	815
Oy	2322	GCGCATCCGCAAGGTGCTGTTCTTGAGACGCAATGATGAGCGGCAATGTGATCTACAGTA	2388
Db	816	CGGCGAAGGGCTCGCGCGTCACTGTCGACGAGATCGACCCCAATCTGCGCGCTGCAAGCGCTC	875
Oy	2382	CATGGAAGACCTGTATACGTGGGCAAGCGGCGGCGCTTAGATCGA	2423
Db	876	GATGGCGGAGTTTCAGATCAACACGCTTGAGGCGGAGCTTCGA	917
RESULT 9			
LOCUS	CL971508	1941 bp	DNA
DEFINITION	OS:FFC021485 Oryza sativa Expressed Library Oryza sativa (indica		
ACCESSION	CU971508		
VERSION	CU971508.1		
KEYWORDS	GSS.		
SOURCE	Oryza sativa (indica cultivar-group)		
ORGANISM	Oryza sativa (indica cultivar-group)		
REFERENCE	Embarayota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.		
AUTHORS	1 (bases 1 to 1941) Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M., Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L., Wong, G. K. S., Deng, X. W. and Wang, J.		
TITLE	An analysis of transcriptional regulation of the rice genome and its comparison to Arabidopsis		
JOURNAL	Unpublished (2004)		
COMMENT	Contact: Chen Chen Department of Bioinformatic Beijing Institute of Genomics Chinese Academy of Sciences, Beijing 101300, China Tel: 86-10-80481559 Fax: 86-10-80488676 Email: chenchen@genomics.org.cn Rice genomic sequence. Class: exon-trapped.		
FEATURES			
source	Location/Qualifiers 1..1941 /organism="Oryza sativa (indica cultivar-group)" /mol_type="genomic DNA" /db_xref="taxon:39946" /clone_lib="Oryza sativa Expressed Library" /note="Oryza sativa exon trapped genomic sequences"		
ORIGIN			
Query Match	3.4%	Score 84.2;	DB 10; Length 1941;
Best local similarity	42.5%	Pred. No. 7.8e-05;	
Matches	705;	Conservative 0;	Mismatches 933; Indels 21; Gaps 4
Oy	1	GTGCAAGCGCACATGCGCGAGCGCATGAGCCAGGCGCACAGCGCCACATCTGATGCG	60
Db	163	GAGCGCGCCCAAGAACAGAGTGCATGAACCAACCAACACACGATCTTGATGCGAAGCGG	222
Oy	61	CGAGCAACTTCAGAGGGGCCCAAGGCGATCATCAAGTGTCTTCACTGGGGCAAGAGGCG	120
Db	223	TTCATGTGTGAGAGAGTTCTTCGACCCGTCGTCGAGATGACATGAACTATAGCGCTTC	282
Oy	121	CACATCGCGCGCAACTGCGCGCGCCCGCGCAAGAAAGGCTGTGGAAGTGGCGCAAGAG	180
Db	283	AAAGTGTGCTCGGCGCCCGGCGCAAGCGCATGATGTCGTCTCAAGTACCAAGGCGAAGAG	342
Oy	181	GCGCACCATGAAGACTGACACCGAGCGCAAGGCAACTTCTTCGCGAAGACTGCGC	240

```

Db      343  AAGCATTTGCGCGGAGAGATTCCTGCTCATGTGCTGATCTAGATGAGGAGATGCGC 402
Qy      241  TTCCCGGAGGCGAGGCGCGGAGATTCCCGAGCGAGAGAACCGCGCCACAGGCCACC 300
Db      403  GAGGCTTACCTCTGGCTGCTCATCAAGAACGCGTGGTCAAGCTCCGCGCTACTTCAAC 462
Qy      301  AGCGCGAGCTGCGAGCTGCGCGGCGAGAACCCCGCGAGCGCGCGCGCGAGCGCCAG 360
Db      463  GACTCGCAGAGGAGGCGAGCAAGAGACCGCGGCTCATCGCGCGGCTCAAGGTATGCGC 522
Qy      361  GGCAACCTTGAATCTTCCCGAGATCAACCTGTGGCGAGCGCGCGCTGTGAGATCAAGTG 420
Db      523  ATCATCAACGAGCCCAACCGCGCGCGCTCATCGGTACGGCTCTGACAAAGAGCGAGCC 582
Qy      421  GCGCGCGAGATCAAGAGAGCGCGCTGTGACACCGCGCGCGAGACACCGCTGTGAGAG 480
Db      583  AGCGCGAGAGAGAGAGTGTCTCATCTTTGACCTCGCGCGCGGAGCGTTGATGTGTGCG 642
Qy      481  ATGAGCTT--GCGCGCGAGTGTGAGAGCCCAAGATGATCGCGCGCATGCGCGCTTCA 537
Db      643  CTCAACATCGAGAGGCGCATCTTCAGAGGTCAAGCGCACCGCGCGAGACCTCACTCG 702
Qy      538  AAGGTGCGCGAGTACGACCAAGATCTGTATCGAGATCTGTGCGGAG--AAGGCCATCG 594
Db      703  GCGGAGGACTTTCGACACCGCATGTGTGAACAATTGTGAGAGAGTTCAAGCGCAAGAAC 762
Qy      595  ACCGCTGTATCGGCGCGCGCGCGCGCGGATCATCTCGCGCGGACATGCTGACCGAG 654
Db      763  AAGAGAGACATCAAGCGGAGACCGCGCGCGCGCTGCGCGCGCTGAGAGCGCGTCCAG 822
Qy      655  GCGTGCACCTTGAATCTTCCCATCAAGCCCATCGAGACCGTGCCTGTGAAGCTGAAG 714
Db      823  GCCAAGCGGAGCGTGTGCTGCGAGCGGAGACCACTATGATGATGATCGCTGTATGAG 882
Qy      715  GGCATGAGACGCGCGCGCGCGCGCGCGCTGTGACCGAGAGAGAGATCAAGCGCTT 774
Db      883  GGCATGAGATCTTCAACGAGATCAACAGGCGCGCGTTCGAGAGGTCAACATGAGACT 942
Qy      775  ACCGCGATCTGAGAGAGATGAGAGAGAGGCGCAAGATCAACAGATCGGCGCGCGAG 834
Db      943  TTCCGAGATGATGAGAGCGCGCGCGCGCGCTGTGAGAGGTCTTCGCGAGCGCAAGT 1002
Qy      835  CCTCAACACACCGCGCTGTTCGCGCATCAAGAGAGAGAGAGAGAGAGAGAGAGAG 894
Db      1003  AGCGTGCAGAGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1062
Qy      895  -----GTGACTTTCGCGAGCTGAACAGCGCAACCGAGACTTCTGTGAGGTG 942
Db      1063  CTCGAGAGATCTTCAACGCGCAAGAGAGCTCTGCAAGAGATCAACCCGACGAGCG 1122
Qy      943  CAGCTGTGATCCCCCGCGCGCGCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1002
Db      1123  GGTACGCGCGCGCGCGCTTCAGCGCGCGCATCTGAGCGCGAGGCGC--AACGAGAG 1179
Qy      1003  GTGGGCGAGCGCTTCAAGCGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1062
Db      1180  CAGGACTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1239
Qy      1063  ACCATCCCGAGATCAACAGAGACCGCGGATCTGCTTCAAGTCAAGTCAAGTCTG 1122
Db      1240  GTCATGACGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 1299
Qy      1123  CAGGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1182
Db      1300  TTCAACTTCTGCAACAGAGACCGCGGCTGTCTTCAAGGTATCAAGAGAGAGAGAG 1359
Qy      1183  TTCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242
Db      1360  AGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1419
Qy      1243  CTGAGAGATCGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1302

```

```

Db      1420  CGCGCGTGCAGATCAAGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1479
Qy      1303  GCGTTACACACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1362
Db      1480  TTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1539
Qy      1363  CACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1422
Db      1540  CGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1599
Qy      1423  AAGCATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1482
Db      1600  GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
Qy      1483  AAGGTGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1542
Db      1660  ATGCGCAACATCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1719
Qy      1543  CTGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1602
Db      1720  AGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1779
Qy      1603  CACGCGGTGTACTAGACCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1641
Db      1780  GCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1818

```

```

RESULT 10
LOCUS   CW922203/c
DEFINITION EDCAR29TR A. castellanii, 6-8 kb library from total genomic DNA
          Acanthamoeba castellanii genomic clone EDCAR29, genomic survey
          sequence.
ACCESSION CW922203
VERSION   CW922203.1
KEYWORDS  GI:60228146
SOURCE    Acanthamoeba castellanii
          Acanthamoeba castellanii
          Eukaryota; Acanthamoebidae; Acanthamoeba.
REFERENCE 1 (bases 1 to 1060)
          Anderson, I.J. and Loftus, B.J.
          Gene discovery in the Acanthamoeba castellanii genome
          JOURNAL Unpublished (2004)
          Contact: Iain Anderson
          The Institute for Genomic Research (TIGR; www.tigr.org)
          9712 Medical Center Drive, Rockville, MD 20850, USA
          Tel: 301-795-7949
          Fax: 301-838-0208
          Class: Shotgun.
FEATURES             Location/Qualifiers
     source            1..1060
                     /organism="Acanthamoeba castellanii"
                     /mol_type="genomic DNA"
                     /strain="Neff"
                     /db_xref="taxon:5755"
                     /clone="EDCAR29"
                     /clone_1ib="A. castellanii, 6-8 kb library from total
                     genomic DNA"
                     /note="Vector: pHD2"

```

ORIGIN

```

Query Match      3.3%; Score 80.8; DB 10; Length 1060;
Best Local Similarity 43.9%; Pred. No. 0.0003;
Matches 392; Conservative 0; Mismatches 497; Indels 3; Gaps 1;

Qy      261  CGAGTTTCCCAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 320
Db      955  CGTGGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 896
Qy      321  CGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 380
Db      895  CGGCTGTGTGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 836

```

QY 381 GATCACCTGTGGCAGCGCCCTCTGTGATGATCAAGGTGGCGGCAGATCAAGAGGC 440
 DB 835 CAAGAGAGTCAAGAGGCGGAGGAGGAGATCTCTTCACTTCAAGATCACTCTGT 776
 QY 441 CTTGTGACACCGGCGGCGGAGACACCGGTGTGAGAGATGAGCTTCCGGGAATG 500
 DB 775 GCTGGGCGCGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 716
 QY 501 GAAGGCCAAGATGATCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 560
 DB 715 CGCGCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 656
 QY 551 CTTGATGAGATCTGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 620
 DB 655 GAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 596
 QY 621 GAAATCTATTCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 680
 DB 595 CGAGGAGACCGGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 536
 QY 681 CCCCATGAGACCGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 740
 DB 535 GCGCATCTATGAGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 476
 QY 741 GTGGCCCTTGAACCGAGAGAGAGATCAAGGCGCTGACCGGCGATCTGCGAGAGAG 797
 DB 475 CGGCTTCTGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 416
 QY 798 GAAAGAGGCGAAGATCAACCAATGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 857
 DB 415 GAGGCTGAGCTGCGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 356
 QY 858 CATCAAG 917
 DB 355 GGTGAGCTCAACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 296
 QY 918 GCGGACCGAGAGCTTCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 977
 DB 295 GCGGATCAAGAGAGATCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 236
 QY 978 GAAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1037
 DB 235 GCGGCGTACAG 176
 QY 1038 GGAATCTGCGAAGTACACCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1097
 DB 175 CGAGGAGCTCAAGGCTCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 116
 QY 1098 CGGCTACCAAGTACAGGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1149
 DB 115 CGATCTGCGGCTACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 64

RESULT 11
 CNI59167/c 869 bp mRNA linear Bst 05-DEC-2003
 LOCUS FGAS040564 Triticum aestivum FGAS: Triticum aestivum cDNA,
 DEFINITION mRNA sequence.
 ACCESSION CNI59167
 VERSION CNI59167.1 GI:38985053
 KEYWORDS Bst.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Bacteria; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 869)
 Allard, F., Crosby, W.L., Danyluk, J., Budes, F., Frick, M., Gaudet, D.,
 Genes, B., Graf, R., Gulick, P., Hryman, L.D., Larocque, A.,
 Links, M.G., McCarthy, E.L., Monroy, A., Muzak, I., Nilsson, D.,
 Pennik, C., Roach, D.L. and Sarhan, F.
 Functional Genomics of Abiotic Stress in Wheat and Canola Crops
 Unpublished (2003)

COMMENT
 Contact: Wm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas@atcc.usask.ca
 This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region 1128,636.
 Plate: Talc537 row: N column: 23.
 Location/Qualifiers
 1..869
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultiVar="wheat line PI 178383"
 /db_xref="taxon:4565"
 /lab_host="DH5 alpha"
 /clone_lib="Triticum aestivum FGAS: Talc5"
 /note="Organ: Crown, Vector: pGEM-T, SSH (suppression
 subtractive hybridization) cDNA library from genotype
 PI178383 cold hardened at 2 C for 21 days and 49 days
 (equal amount of cDNA pooled together before subtraction,
 tester) and subtracted against genotype Norstar cold
 hardened at 2 C for 1 day (24 H) (driver). Modified Smart
 cDNA (Clontech) priming and non-directional cloning"

ORIGIN
 Query Match 3.3%; Score 80.6; DB 7; Length 869;
 Best Local Similarity 44.9%; Pred. No. 0.00033;
 Matches 305; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

QY 580 AAGAGGCGATGCGGAGCGGTGATGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 639
 DB 827 ATGGCGGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 768
 QY 640 ATGCTGAGCCGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 699
 DB 767 CACAAAG 708
 QY 700 GTGAAGCTGAAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 759
 DB 707 ACGAAG 648
 QY 760 AAGATCAAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 819
 DB 647 AACAAAG 588
 QY 820 ATGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 879
 DB 587 AACAAAG 528
 QY 880 AAGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 939
 DB 527 AACAAAG 468
 QY 940 GTGAGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 999
 DB 467 AACAAAG 408
 QY 1000 GAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1059
 DB 407 GACAAAG 348
 QY 1060 TTCACATCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1119
 DB 347 GACCAAG 288
 QY 1120 CCCAGGCGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1179

[illegible]

Db	125	CTACACAGTCAAGACGACGTCAACACGACGTTAACACAGACATCGACGTGCACAC	184
OY	1824	GTTCCGCTCGCCATTCAGAAAGAGACCTGGAGACCTGTGACCGACTACCTGGACGCG	1883
Db	185	CAMGCAACACTACGTCACACCACTTCGACGACACCGACGACTACCTGAGCTGACGCT	244
OY	1884	CACCTGATCCCGAAGTGGGAGTTCTGGAAACCCCCCTGTGTAAAGCTGTGTACCA	1943
Db	245	CAAGCAACAGTCAACGACGCTCAACGACACTTGACGTTCAGACACCAAGCCCAA	304
OY	1944	GCTGAAAGAGAGCCATTCATCGGCGCGGAGACCTTCTACGTGACCGGCGCCCAACG	2003
Db	305	CCAGTGCACAAACACTTTCAGACGACACCGACGACTTACTCTCGAGTACCTCGACGACGA	364
OY	2004	CGAGACCAAGATCGGCAAGGCGGCTTACGTGACCGACCGGGGCGGCGAAGATCGTAG	2063
Db	365	CGTCAACAGAGACTTCGACGTGTGACGACCAAGACCAACAGTCCACCAACACTTCGA	424
OY	2064	CTTGAACGAGACCAACCA---CCAGAAACCGAGCTGACAGGCCATTCAGCTGTGGCCCTGCA	2120
Db	425	CGACACCGACGACTACTGCTCGAGCAGCTTCAGACACAGTCAACGACGACGCTCAACCA	484
OY	2121	GGAAGGCGGACGAGGTTGAAACATCTGACCGACGACGACGATGACGCTGTGGCATCATCCA	2180
Db	485	GCACTTCGAGGTGCACGACACCAAGACCAACACAGTCCACCAACACTTCGACGACACCGA	544
OY	2181	GGCCCAAGCCGACCAAGACGACGAGCGAGCTGTGTAAACAGATTCATTCGACGACTGATCAA	2240
Db	545	CGACTTACTAGTTCGAGCACGTCAACACGACAGTCAACAGACGTCGACGACGACCATTCGA	604
OY	2241	GAAGAGAAAGTGTACTTGAGCTGTGGTGGCCGCCCAAGAGGCATTCGGCGGCAACGACGA	2300
Db	605	CGTGCAGACACCAAGACCAACCACTGTCCAAACCACTGGCGACGACGCGACGCTCTCTA	664
OY	2301	GATGACAGAGTGTGTGAGCAAGGGCATTCGCAAGTGTCTGTCTCTGACCGGCATTCGATG	2359
Db	665	GCTGG---AGACCTTCAACGACGACGTCACACGACACTTCGACGTTCGACGACACCGAGG	720

FEATURES	source
LOCUS	BM320864
DEFINITION	rockefeller.0.46 Masticamoeba balamuthi lambda ZAP II library
ACCESSION	BM320864
VERSION	BM320864.1
KEYWORDS	GI:18055270
SOURCE	EST.
ORGANISM	Masticamoeba balamuthi
REFERENCE	Masticamoeba balamuthi
AUTHORS	Eukaryota: Pelobiontida, Masticamoebidae; Masticamoeba. 1 (bases 1 to 1132)
	Bapteste, S., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., Durrille, L., Gaasterland, T., Lopez, P., Muller, M., and Philippe, H.
	The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Hartmannella, and Masticamoeba Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
COMMENT	11830664
JOURNAL	Contact: Muller Miklos
PUBMED	Laboratory of Biochemical Parasitology The Rockefeller University 1230 York Avenue, New York, NY 10021, USA Email: mmuller@rockvax.rockefeller.edu Insert Length: 1132 Std Error: 0.00 POLYA=No.
	location/Qualifiers
	1. 1132
	/organism="Masticamoeba balamuthi"
	/mol_type="mRNA"
	/strain="ATCC 30984"
	/db_xref="taxon:108607"
	/clone_id="Masticamoeba balamuthi lambda ZAP II library"

ORIGIN /note="syn: Phreatamoeba balamuthi"

Query Match 3.3%; Score 80.2; DB 3; Length 1132;
Best Local Similarity 45.3%; Pred. No. 0.0039;
Matches 375; Conservative 0; Mismatches 444; Indels 9; Gaps 2;

QY CGCGCCCCCGCAAGAGAGGCTGCTGGAAGTGCAGCAAGAGAGGCGCACAGTGAAGAC 198
DB CGCGCGAGAGGAG 138
QY TGCAACGAGCGCCAGAGGCACTTCTTCCGAGAGAGCTGGCGCTTCCCGAGGCAAGGCC 258
DB TACAACAGCGCCAG 192
QY CGCGAGTTCGCGAG 318
DB TGCGAGATCGCGCTTAAG 252
QY CGCGGAG 378
DB GAGCTACCGCGCTTGGCGCTGAAGTCCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 312
QY CAGATCACCGCTGAG 438
DB GAGCGTGTGCTGAG 372
QY GCGCTGTGAG 498
DB GTCAAG 432
QY TGAAG 558
DB TTCAAG 492
QY ATCTGATGAG 618
DB GCCCTCAAG 549
QY GTGAACATCATGAG 678
DB GTGAGCTTCAAG 609
QY AGCCCATGAG 738
DB GCGCGGAG 669
QY CAGTGGCGCGTGAAG 798
DB CGCGAGTTCTGCGCTAG 729
QY AAG 858
DB ACCGAG 789
QY ATCAAG 918
DB CCCGAG 849
QY CGCAGCAG 966
DB AACCGAG 897

RESULT 14
BM320900 1165 bp mRNA linear EST 03-JAN-2002
LOCUS rocheffeller.0.353 Mastigamoeba balamuthi lambda ZAP II library
DEFINITION Mastigamoeba balamuthi cDNA similar to ribosomal protein L5, mRNA
ACCESSION BM320900
VERSION BM320900.1 GI:18055306
KEYWORDS EST.

SOURCE
ORGANISM Mastigamoeba balamuthi
Mastigamoeba balamuthi
Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.

REFERENCE
AUTHORS 1 (bases 1 to 1165)
Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sengen, C.W.,
Gordon, P., Durrille, L., Gaasterland, T., Lopez, P., Muller, M. and
Philippe, H.

TITLE
JOURNAL The analysis of 100 genes supports the grouping of three highly
PUBMED divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
11830664 Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)

COMMENT
Contact: Muller Miklos
Laboratory of Biochemical Parasitology
The Rockefeller University
1230 York Avenue, New York, NY 10021, USA
Email: mmuller@rockefeller.edu
Insert Length: 1165 Std Error: 0.00
POLYA=No.

FEATURES
Location/Qualifiers
1..1165

source
/organism="Mastigamoeba balamuthi"
/mol_type="mRNA"
/strain="ATCC 30984"
/db_xref="taxon:108607"
/clone_id="Mastigamoeba balamuthi lambda ZAP II library"
/note="syn: Phreatamoeba balamuthi"

ORIGIN

Query Match 3.2%; Score 79.6; DB 3; Length 1165;
Best Local Similarity 45.1%; Pred. No. 0.0005;
Matches 428; Conservative 0; Mismatches 506; Indels 16; Gaps 3;

QY CATGAGTCTTCAAGTGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 149
DB CGTCAAG 105
QY CAAGAAG 209
DB CAAGAAG 165
QY CGAGGCACTTCTTCCGAG 269
DB CAAGTACCGCTTCTGTC-----GTCCGCTTCAAGAGAGAGAGAGAGAGAGAGAGAG 219
QY CAGGAG 329
DB CTAGGCAAG 279
QY CCGCGAG 389
DB CTTGAGGCTCAAGTCTGAG 339
QY GTGAG 449
DB GCGCGGAG 399
QY CACGAG 509
DB CAAGGAG 457
QY GATGATGAG 569
DB TGTCTGAG 515
QY GATCTGCGAG 629
DB GGGAGATGAG 575
QY CGGCGGAG 689
DB CGGCGAG 635
QY GACGCTGCGCTGAAG 749

Db 636 CGCCGCTGATGATGAGCTCTCTAAGAGAGAGACCCGCGCTTTCGACCCGCAATTCTC 695
Qy 750 GACCGAGAGAGAGATCAAGGCGCTTGAACCGCATCTGCGAGAGATGAGAGAGAGCGCA 809
Db 696 GCGCTAACGCAAGAGAGGCTGTCAACGCGCGACATGCTCGAGAGATCTACACGAGGCCCA 755
Qy 810 GATCACAAGATCGGCGCCCGAGAACCCCTTACAAACCCCGTGTTCGCTCATCAAGAGAA 869
Db 756 -----CAAGCAGATTCGCGCGCGACCGCACTTTCGTCCTCCAGCGCGCTCGAACCCGA 809
Qy 870 GGACAGACCAAGTGGCGCGAGCTGTGAGCTTTCGCGAGCTGAACAAGCCACCGAGA 929
Db 810 GGGCGCCAGGCCCAAGCACTTGGGGCAAGCGCTGACGTACACAGAGCGCAAGAACCG 869
Qy 930 CTTCGTGGAGTGTGACGTGGGCGATCCCGACCCCGCGCGCTGAGAGAGAGAGCGCT 989
Db 870 CGTCCGCCAGAGAGAGGTCCGCGTGGGCTAACCCCGACCGCCCAAGAGCGCAATAATTCC 929
Qy 990 GACCGTCTGTGACGTGGGCGAGCGCTTACTTCAAGCGTGGCCCTGAGACGAGG 1039
Db 930 CGGTCCCTGTGACACCGCGCATCTCCGCTGCGCGTGTGCGCTGTGCGG 979

RESULT 15
CL982770 1680 bp DNA linear GSS 21-SEP-2004
LOCUS OaIRSC049024 Oryza sativa Expressed library Oryza sativa (indica
DEFINITION cultivar-group) genomic, genomic survey sequence.
ACCESSION CL982770
VERSION CL982770.1 GI:52420015
KEYWORDS GSS.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group) Embryophyta; Tracheophyta;
Eukaryota; Viridiplantae; Streptophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 1680)
Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
Wong, G. K. S., Deng, X. W. and Wang, J.
An analysis of transcriptional regulation of the rice genome and
its comparison to Arabidopsis
JOURNAL Unpublished (2004)
COMMENT Contact: Chen Chen
Department of Bioinformatic
Beijing Institute of Genomics
Chinese Academy of Sciences, Beijing 101300, China
Tel: 86-10-80481559
Fax: 86-10-80488676
Email: chenchen@genomics.org.cn
Rice genomic sequence.
Class: exon-trapped.

FEATURES
source Location/Qualifiers
1..1680
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="genomic DNA"
/db_xref="taxon:39946"
/clone_lib="Oryza sativa Expressed library"
/note="Oryza sativa exon trapped genomic sequences"

ORIGIN
Query Match 3.2%; Score 79.6; DB 10; Length 1680;
Best Local Similarity 41.9%; Pred. No. 0.00051;
Matches 612; Conservative 0; Mismatches 844; Indels 6; Gaps 2;

Qy 270 CAGCGAGCAGAACCGCGCCCAACAGCCCGACAGCGCGAGCTGACAGTGGCGGCGCA 329
Db 129 CATCAACGTGAGACCGCGGCAACATGTCTCCGTTCCAGCAAGCGAGTGTTCGTACCGA 188
Qy 330 CCCCCGAGAGAGCGCGCGCGAGCGCGAGCGCGACCTTGAATTTCCCGACATCACCTT 389
Db 189 CAGCGCGCGCGAGACCGACCTCGACTCTCGGCGCACTAGAGCGTTTATCGAAGACGCGCAT 248

Qy 390 GTGGAGCGCCCGCTGTGAGCATCAAGTGGCGGCGGAGATCAAGAGGCGCTGTGGA 449
Db 249 GAAGAGACCAACAACTTACACACGCGCGCATCTTACAGTCTCGTGAAGAGAGAGCG 308
Qy 450 CACCGCGCGCGACGACACCGTGTGAGAGAGATGAGCTGCGCGCGCAAGTGAAGCCCA 509
Db 309 CCGCGCGCATCTACCTCGCGCAAGACGCTGTGAGATATCCGACATCTACCAAGAGATCCA 368
Qy 510 GATGATCGGC--GCGATCGCGCGCTTATCAAGTGGCGGCAATGACAGATCTGAT 566
Db 369 GGAATATGTCAGAGCGCGCGCGGACATCGGACCGACAGCGCGGTGAGCTGTGCATCGT 428
Qy 567 CGAATCTGCGCGCAAGAGGCGCATGCGACCGTGTGATCGGCGCCACCCCGTGAACAT 626
Db 429 CGAGATCGCGCGCACCGTGGCGGAGATGGAATGCTGCGTTCCTGAGAGCGGTGCGCA 488
Qy 627 CATCGCGCGCAACATGTGACCCAGCTGGGCTGACCCCTGAACCTTCCCATGAGCCCAT 686
Db 489 GATGAGCTTGGCGCATGCGGCGGACAAATTCGCGCTTCTGTGACCTTACCTTACCTG 548
Qy 687 CGAGACCGGTGCGCTGAGAGCTGAGAGCCGCGCATGAGCGGCGCCAGAGTGAAGAGTGGCC 746
Db 549 CATCGCGCGCGCGGTGAATCTAAGACCAAGCCACCGACGACAGCTGTGAGAGAGCTGCG 608
Qy 747 CTTGACCGAGAGAGAGATCAAGGCGCTTGAACCGCCATCTGTGAGAGATGAGAGAGAGG 806
Db 609 CGAGATTCGCGCATTCAGCGCGCGCTTCTGTGCGCGCGCGACCGCGCGATCCGACGA 668
Qy 807 CAAGATCAACAAGATCGGCGCGCGCGAGACCCCTTACAAACCCCGGTTCGCGCATGAAGA 866
Db 669 GAGGCGCGCGAGATCTCGCTGTTCACAACTGCGCGCAATGAGGCGGTGATGAGATATG 728
Qy 867 GAAAGCAGACCAAGTGGCGCGAGCTGTGAGCTTTCGCGAGCTGAACAAGCGACCCA 926
Db 729 GAGCGTGAACACCATCTTCAAGGTGCGCGCGCATGTGACGAGCGAGGCGTGAAGCGCT 788
Qy 927 GACCTTCTGGAGAGTGTGACGTGGGATATCCCGACCCCGCGCGCTGAGAGAGAGAG 986
Db 789 GATCTGCGCAAGAGCTGCGCGCTGGAACCGCGCGCACAGCTTCMAAGCGCTGAGACCACT 848
Qy 987 CGTGAACGCTGTGAGCGTGGGCGAGCGCTTACCTTACGCGTGGCGCTGAGACGAGACTTCCG 1046
Db 849 CGTGAACGAGACCGGAGCATTCGCGAGGCGAGGCTCACATCGCATGTGGCGCAATGACT 908
Qy 1047 CAAGTACACCGGCTTACCATTCGCCAGCATCAACAGAGAACCCCGCGCATCCGTAACA 1106
Db 909 AGACCTGTCCAGACACTTCAAGTGGTGAACGAGGCGGTGCGCGCACGCGCGCATGAAGA 968
Qy 1107 GTACAACTGTGCTGCGCGCGAGGCTGGAAGGCGAGCGCCAGCATCTTTCAGAGCAGATGAC 1166
Db 969 CCACGTGCGCGTGAAGATCGAGACAGCGTCAATTCAGAGACATTCGCGCGCGAGCGCGC 1028
Qy 1167 CAAGATCTGTGAGAGCGCTTCCGCGCGCGCGCAACCCGAGATGTGATCTACAGGCGCCCT 1226
Db 1029 CGAGAGCTGTGCGAGGTGAGATGATGCCATCTGTGTGCGCGCGCGCTTGTGCGCGCGCT 1088
Qy 1227 GTACGTGGCGAGCGACTTGAAGATCGCGAGACCGCGCGCAAGATCGAGAGCTGTGCA 1286
Db 1089 GGAAGGCAAGATCTGCAAGCGCGCGGTGAGCGCGCGAGAGAGGTGCGCTTACTGTGGCAT 1148
Qy 1287 GCACTGTCTGCGCGTGGGCTTACCAACCCCGCAAGAGACACAGAGAGAGAGAGAGAG 1346
Db 1149 CTGCTGTGGCATGAGTGGCGGACCATGAAATATGCGCGCGCATGTGGCGCGCTGCGGA 1208
Qy 1347 CCGTCCCATGTGAGCTGACCCCGACAGTGAACCGTGAAGCGCCCATGAGCTGCGCGAGA 1406
Db 1209 CGCGAAGCAGACCGAATTCATGATCGCGCGCACGCGCGCATTCGTAATGCGCTATACCGA 1268
Qy 1407 GGAAGCTGGAACCGTGAACGACATTCAGAGTGTGTGGCAAGTGAAGTGGCGCGAGCCA 1466
Db 1269 GTGAGAGAGCGCGGACCGGACCATTCAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1328

QY 1467 GATCATCCCGGACATCAAGGTGGCGCACTGTGCAAGCTGCTGGCGCGCGCAAGGCTT 1526

Db 1339 CATGGCCCTTGGGCGGGCAAGCTTCCGAGCTGACCTCCGGCACGCTGGGGCGCAGAAATTTA 1388

QY 1577 GACCGACATCTGTGCTCCCTTGACCCGAGAGAGCGAGCTGGAGCTGGCCGAGAAACCGCGAGAT 1586

Db 1389 CGGCAACATCTGTCAACCGAGCGCCACCGCCACCGGTATCCAGGCGCAAGCTGAATCTACTTGA 1448

QY 1587 CCTGGCGCAGGCGCGGTGCAACGGCGTGTACTCAAGACCCGACGAAG---GACCTGTGGTGGCCGA 1643

Db 1449 CCAAGTCCCGCAAGGGCGGGCTCTGTATCTCCGCGCTCAACGACGGCGCAACAGCTCACCGGA 1508

QY 1644 GATCCAGAAGCAGGGCCACAGCAAGTGAACCTTACCAAGATTTAACGAGAGCCCTTCAAGAA 1703

Db 1509 GATCTCTGAGCTGCCGCGCGAGAGTCCATCCGTGTTCACTGGCGGTGCAAGTTCCACCCGGA 1568

QY 1704 CCTGAAGAACGGCAGAGTACGCC 1725

Db 1569 GTTCAAGTCCACCCCGTGGGAC 1590

Search completed: December 31, 2005, 02:31:32
Job time : 8682.48 secs